Stochastic Processes and Financial Mathematics (part one)

Dr Nic Freeman / Dr Robin Stephenson

September 4, 2024

Contents

Chapter 0

Introduction

We live in a random world: we cannot be certain of tomorrow's weather or what the price of petrol will be next year – but randomness is never 'completely' random. Often we know, or rather, believe that some events are likely and others are unlikely. We might think that two events are both possible, but are unlikely to occur together, and so on.

How should we handle this situation? Naturally, we would like to understand the world around us and, when possible, to anticipate what might happen in the future. This necessitates that we study the variety of random processes that we find around us.

We will see many and varied examples of random processes throughout this course, although we will tend to call them *stochastic processes* (with the same meaning). They reflect the wide variety of unpredictable ways in which reality behaves. We will also introduce a key concept used in the study of stochastic processes, known as a martingale.

It has become common, in both science and industry, to use highly complex models of the world around us. Such models cannot be magicked out of thin air. In fact, in much the same way as we might build a miniature space station out of individual pieces of Lego, what is required is a set of useful pieces that can be fitted together into realistic models. The theory of stochastic processes provides some of the most useful building blocks, and the models built from them are generally called stochastic models.

One industry that makes extensive use of stochastic modelling is *finance*. In this course, we will often use financial models to motivate and exemplify our discussion of stochastic processes.

The central question in a financial model is usually how much a particular object is worth. For example, we might ask how much we need to pay today, to have a barrel of oil delivered in six months time. We might ask for something more complicated: how much would it cost to have the opportunity, in six months time, to buy a barrel of oil, for a price that is agreed on today? We will study the Black-Scholes model and the concept of 'arbitrage free pricing', which provide somewhat surprising answers to this type of question.

0.1 Organization

Syllabus

These notes are for two courses: MAS352 and MAS61023. This is part one of the lecture notes. Part two will be made available when we reach Chapter 10.

Some sections of the course are included in MAS61023 but not in MAS352. These sections are marked with a (Δ) symbol. We will not cover these sections in lectures. Students taking MAS61023 should study these sections independently.

Some parts of the notes are marked with a ($\oslash$) symbol, which means they are off-syllabus. These are often cases where detailed connections can be made to and from other parts of mathematics.

Problem sheets

The exercises are divided up according to the chapters of the course. Some exercises are marked as 'challenge questions' – these are intended to offer a serious, time consuming challenge to the best students.

Aside from challenge questions, it is expected that students will attempt all exercises (for the version of the course they are taking) and review their own solutions using the typed solutions provided at the end of these notes, in Appendices A and C.

At three points during each semester, an assignment of additional exercises will be set. About one week later, a mark scheme will be posted, and you should self-mark your solutions.

Examination

Both versions of the course will be examined in the summer sitting. Parts of the course marked with a (Δ) are examinable for MAS61023 but not for MAS352. Parts of the course marked with a ($\oslash$) will not be examined (for everyone).

A formula sheet will be provided in the exam, see Appendices B (for semester 1) and E (for semester 2). Some detailed advice on revision can be found in Appendix D, attached to the second semester notes.

MAS61023 also contains a mid-year online test, which will take place towards the end of January and comprise 15% of the final mark. MAS352 is assessed entirely by exam.

Website

Further information, including the timetable, can be found on

`https://nicfreeman1209.github.io/Website/MASx52/`.

Chapter 1

Expectation and Arbitrage

In this chapter we look at our first example of a financial market. We introduce the idea of arbitrage free pricing, and discuss what tools we would need to build better models.

1.1 Betting on coin tosses

We begin by looking at a simple betting game. Someone tosses a fair coin. They offer to pay you \$1 if the coin comes up heads and nothing if the coin comes up tails. How much are you prepared to pay to play the game?

One way that you might answer this question is to look at the expected return of playing the game. If the (random) amount of money that you win is X, then you'd expect to make

$$\mathbb{E}[X] = \tfrac{1}{2}\$1 + \tfrac{1}{2}\$0 = \$0.50.$$

So you might offer to pay \$0.50 to play the game.

We can think of a single play as us paying some amount to buy *a random quantity*. That is, we pay \$0.50 to buy the random quantity X, then later on we discover if X is \$1 or \$0.

We can link this 'pricing by expectation' to the long term average of our winnings, if we played the game multiple times. Formally this uses the strong law of large numbers:

Theorem 1.1.1 *Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of random variables that are independent and identically distributed. Suppose that $\mathbb{E}[X_1] = \mu$ and $\mathrm{var}(X_1) < \infty$, and set*

$$S_n = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Then, with probability one, $S_n \to \mu$ as $n \to \infty$.

In our case, if we played the game a large number n of times, and on play i our winnings were X_i, then our average winnings would be $S_n \approx \mathbb{E}[X_1] = \frac{1}{2}$. So we might regard \$0.50 as a 'fair' price to pay for a single play. If we paid less, in the long run we'd make money, and if we paid more, in the long run we'd lose money.

Often, though, you might not be willing to pay this price. Suppose your life savings were \$20,000. You probably (hopefully) wouldn't gamble it on the toss of a single coin, where you would get \$40,000 on heads and \$0 on tails; it's too risky.

It is tempting to hope that the fairest way to price anything is to calculate its expected value, and then charge that much. As we will explain in the rest of Chapter 1, this tempting idea turns out to be *completely* wrong.

1.2 The one-period market

Let's replace our betting game by a more realistic situation. This will require us to define some terminology. Our convention, for the whole of the course, is that when we introduce a new piece of financial terminology we'll write it in **bold**.

A **market** is any setting where it is possible to buy and sell one or more quantities. An object that can be bought and sold is called a **commodity**. For our purposes, we will always define exactly what can be bought and sold, and how the value of each commodity changes with time. We use the variable t for time.

It is important to realize that money itself is a commodity. It can be bought and sold, in the sense that it is common to exchange money for some other commodity. For example, we might exchange some money for a new pair of shoes; at that same instant someone else is exchanging a pair of shoes for money. When we think of money as a commodity we will usually refer to it as **cash** or as a cash bond.

In this section we define a market, which we'll then study for the rest of Chapter 1. It will be a simple market, with only two commodities. Naturally, we have plans to study more sophisticated examples, but we should start small!

Unlike our coin toss, in our market we will have *time*. As time passes money earns interest, or if it is money that we owe we will be required to pay interest. We'll have just one step of time in our simple market. That is, we'll have time $t = 0$ and time $t = 1$. For this reason, we will call our market the **one-period market**.

Let $r > 0$ be a deterministic constant, known as the **interest rate**. If we put an amount x of cash into the bank at time $t = 0$ and leave it there until time $t = 1$, the bank will then contain

$$x(1 + r)$$

in cash. The same formula applies if x is negative. This corresponds to borrowing C_0 from the bank (i.e. taking out a loan) and the bank then requires us to pay interest on the loan.

Our market contains cash, as its first commodity. As its second, we will have a **stock**. Let us take a brief detour and explain what is meant by stock.

Firstly, we should realize that companies can be (and frequently are) owned by more than one person at any given time. Secondly, the 'right of ownership of a company' can be broken down into several different rights, such as:

- The right to a share of the profits.
- The right to vote on decisions concerning the companies future – for example, on a possible merger with another company.

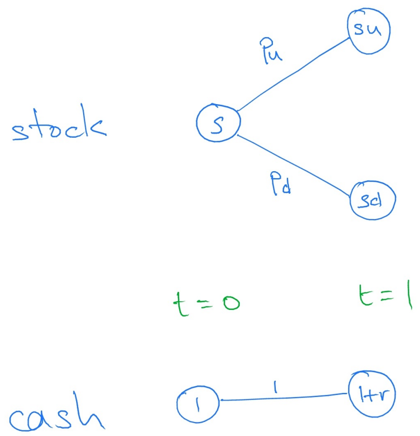
A **share** is a proportion of the rights of ownership of a company; for example a company might split its rights of ownership into 100 equal shares, which can then be bought and sold individually. The value of a share will vary over time, often according to how the successful the company is. A collection of one or more shares in a company is known as stock.

We allow the amount of stock that we own to be any real number. This means we can own a fractional amount of stock, or even a negative amount of stock. This is realistic: in the same way as we could borrow cash from a bank, we can borrow stock from a stockbroker! We don't pay any interest on borrowed stock, we just have to eventually return it. (In reality the stockbroker would charge us a fee but we'll pretend they don't, for simplicity.)

The **value** or **worth** of a stock (or, indeed any commodity) is the amount of cash required to buy a single unit of stock. This changes, randomly:

> Let $u > d > 0$ and $s > 0$ be deterministic constants. At time $t = 0$, it costs $S_0 = s$ cash to buy one unit of stock. At time $t = 1$, one unit of stock becomes worth
>
> $$S_1 = \begin{cases} su & \text{with probability } p_u, \\ sd & \text{with probability } p_d. \end{cases}$$
>
> of cash. Here, $p_u, p_d > 0$ and $p_u + p_d = 1$.

We can represent the changes in value of cash and stocks as a tree, where each edge is labelled by the probability of occurrence.



To sum up, in the one-period market it is possible to trade stocks and cash. There are two points in time, $t = 0$ and $t = 1$.

- If we have x units of cash at time $t = 0$, they will become worth $x(1 + r)$ at time $t = 1$.
- If we have y units of stock, that are worth $yS_0 = sy$ at time $t = 0$, they will become worth

  $$yS_1 = \begin{cases} ysu & \text{with probability } p_u, \\ ysd & \text{with probability } p_d. \end{cases}$$

  at time $t = 1$.

We place no limit on how much, or how little, of each can be traded. That is, we assume the bank will loan/save as much cash as we want, and that we are able to buy/sell unlimited amounts of stock at the current market price. A market that satisfies this assumption is known as **liquid** market.

For example, suppose that $r > 0$ and $s > 0$ are given, and that $u = \frac{3}{2}$, $d = \frac{1}{3}$ and $p_u = p_d = \frac{1}{2}$. At time $t = 0$ we hold a portfolio of 5 cash and 8 stock. What is the expected value of this portfolio at time $t = 1$?

Our 5 units of cash become worth $5(1+r)$ at time 1. Our 8 units of stock, which are worth $8S_0$ at time 0, become worth $8S_1$ at time 1. So, at time $t = 1$ our portfolio is worth $V_1 = 5(1+r)+8S_1$ and the expected value of our portfolio and time t is

$$\begin{aligned}\mathbb{E}[V_1] &= 5(1+r) + 8sup_u + 8sdp_d \\ &= 5(1+r) + 6s + \frac{4}{3}s \\ &= 5 + 5r + \frac{22}{3}s.\end{aligned}$$

1.3 Arbitrage

We now introduce a key concept in mathematical finance, known as **arbitrage**. We say that arbitrage occurs in a market if it possible to make money, for free, without risk of losing money.

There is a subtle distinction to be made here. We might sometimes *expect* to make money, on average. But an arbitrage possibility only occurs when it is possible to make money without any chance of losing it.

Example 1.3.1 Suppose that, in the one-period market, someone offered to sell us a single unit of stock for a special price $\frac{s}{2}$ at time 0. We could then:

1. Take out a loan of $\frac{s}{2}$ from the bank.
2. Buy the stock, at the special price, for $\frac{s}{2}$ cash.
3. Sell the stock, at the market rate, for s cash.
4. Repay our loan of $\frac{s}{2}$ to the bank (we still are at $t = 0$, so no interest is due).
5. Profit!

We now have no debts and $\frac{s}{2}$ cash, with certainty. This is an example of arbitrage.

Example 1.3.1 is obviously artificial. It does illustrates an important point: no one should sell anything at a price that makes an arbitrage possible. However, if nothing is sold at a price that would permit arbitrage then, equally, nothing can be bought for a price that would permit arbitrage. With this in mind:

We assume that no arbitrage can occur in our market.

Let us step back and ask a natural question, about our market. Suppose we wish to have a single unit of stock delivered to us at time $T = 1$, but we want to agree in advance, at time 0, what price K we will pay for it. To do so, we would enter into a **contract**. A contract is an agreement between two (or more) parties (i.e. people, companies, institutions, etc) that they will do something together.

Consider a contract that refers to one party as the buyer and another party as the seller. The contract specifies that:

At time 1, the seller will be paid K cash and will deliver one unit of stock to the buyer.

A contract of this form is known as a **forward** contract. Note that no money changes hands at time 0. The price K that is paid at time 1 is known as the **strike price**. The question is: what should be the value of K?

In fact, there is *only one* possible value for K. This value is

$$K = s(1+r). \tag{1.1}$$

Let us now explain why. We argue by contradiction.

- Suppose that a price $K > s(1+r)$ was agreed. Then we could do the following:

  1. At time 0, enter into a forward contract as the seller.
  2. Borrow s from the bank, and use it buy a single unit of stock.
  3. Wait until time 1.
  4. Sell the stock (as agreed in our contract) in return for K cash.
  5. We owe the bank $s(1+r)$ to pay back our loan, so we pay this amount to the bank.
  6. We are left with $K - s(1+r) > 0$ profit, in cash.

  With this strategy we are *certain* to make a profit. This is arbitrage!

- Suppose, instead, that a price $K < s(1+r)$ was agreed. Then we could:

  1. At time 0, enter into a forward contract as the buyer.
  2. Borrow a single unit of stock from the stockbroker.
  3. Sell this stock, in return for s cash.
  4. Wait until time 1.
  5. We now have $s(1+r)$ in cash. Since $K < s(1+r)$ we can use K of this cash to buy a single unit of stock (as agreed in our contract).
  6. Use the stock we just bought to pay back the stockbroker.
  7. We are left with $s(1+r) - K > 0$ profit, in cash.

  Once again, with this strategy we are *certain* to make a profit. Arbitrage!

Therefore, we reach the surprising conclusion that the only possible choice is $K = s(1+r)$. We refer to $s(1+r)$ as the arbitrage free value for K. This is our first example of an important principle:

> The absence of arbitrage can force prices to take particular values. This is known as **arbitrage free pricing**.

Expectation versus arbitrage

What of pricing by expectation? Let us, temporarily, forget about arbitrage and try to use pricing by expectation to find K.

The value of our forward contract at time 1, from the point of view of the buyer, is $S_1 - K$. It costs nothing to enter into the forward contract, so if we believed that we should price the contract

Figure 1.1: The stock price in GBP of Lloyds Banking Group, from September 2011 to September 2016.

by its expectation, we would want it to cost nothing! This would mean that $\mathbb{E}[S_1 - K] = 0$, which means we'd choose

$$K = \mathbb{E}[S_1] = sup_u + sdp_d. \tag{1.2}$$

This is *not* the same as the formula $K = s(1+r)$, which we deduced in the previous section.

It is possible that our two candidate formulas for K 'accidentally' agree, that is $up_u + dp_d = s(1+r)$, but they are only equal for very specific values of u, p_u, d, p_d, s and r. Observations of real markets show that this doesn't happen.

It may feel very surprising that (1.2) is different to (1.1). The reality is, that financial markets are arbitrage free, and the *correct* strike price for our forward contract is $K = s(1+r)$. However intuitively appealing it might seem to price by expected value, it is not what happens in reality.

Does this mean that, with the correct strike price $K = s(1+r)$, on average we either make or lose money by entering into a forward contract? Yes, it does. But investors are often not concerned with average payoffs – the world changes too quickly to make use of them. Investors are concerned with what happens to them *personally*. Having realized this, we can give a short explanation, in economic terms, of why markets are arbitrage free.

If it is possible to carry out arbitrage within a market, traders will[1] discover how and immediately do so. This creates high demand to buy undervalued commodities. It also creates high demand to borrow overvalued commodities. In turn, this demand causes the price of commodities to adjust, until it is no longer possible to carry out arbitrage. The result is that the market constantly adjusts, and stays in an equilibrium in which no arbitrage is possible.

Of course, in many respects our market is an imperfect model. We will discuss its shortcomings, as well as produce better models, as part of the course.

Remark 1.3.2 We will not mention 'pricing by expectation' again in the course. In a liquid market, arbitrage free pricing is what matters.

[1] Usually.

1.4 Modelling discussion

Our proof that the arbitrage free value for K was $s(1+r)$ is mathematically correct, but it is not ideal. We relied on discovering specific trading strategies that (eventually) resulted in arbitrage. If we tried to price a more complicated contract, we might fail to find the right trading strategies and hence fail to find the right prices. In real markets, trading complicated contracts is common.

Happily, this is precisely the type of situation where mathematics can help. What is needed is a *systematic* way of calculating arbitrage free prices, that *always* works. In order to find one, we'll need to first develop several key concepts from probability theory. More precisely:

- **We need to be able to express the idea that, as time passes, we gain information.**

  For example, in our market, at time $t = 0$ we don't know how the stock price will change. But at time $t = 1$, it has already changed and we do know. Of course, real markets have more than one time step, and we only gain information gradually.

- **We need stochastic processes.**

  Our stock price process $S_0 \mapsto S_1$, with its two branches, is too simplistic. Real stock prices have a 'jagged' appearance (see Figure 1.1). What we need is a library of useful stochastic processes, to build models out of.

In fact, these two requirements are common to almost all stochastic modelling. For this reason, we'll develop our probabilistic tools based on a wide range of examples. We'll return to study (exclusively) financial markets in Chapter 5, and again in Chapters 15-19.

1.5 Exercises on Chapter 1

On the one-period market

All these questions refer to the market defined in Section 1.2 and use notation u, d, p_u, p_d, r, s from that section.

1.1 Suppose that our portfolio at time 0 has 10 units of cash and 5 units of stock. What is the value of this portfolio at time 1?

1.2 Suppose that $0 < d < 1 + r < u$. Our portfolio at time 0 has $x \geq 0$ units of cash and $y \geq 0$ units of stock, but we will have a debt to pay at time 1 of $K > 0$ units of cash.

(a) Assuming that we don't buy or sell anything at time 0, under what conditions on x, y, K can we be certain of paying off our debt?

(b) Suppose that do allow ourselves to trade cash and stocks at time 0. What strategy gives us the best chance of being able to pay off our debt?

1.3 (a) Suppose that $0 < 1 + r < d < u$. Find a trading strategy that results in an arbitrage.

(b) Suppose instead that $0 < d < u < 1 + r$. Find a trading strategy that results in an arbitrage.

Revision of probability and analysis

1.4 Let Y be an exponential random variable with parameter $\lambda > 0$. That is, the probability density function of Y is

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Calculate $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$. Hence, show that $\text{var}(X) = \frac{1}{\lambda^2}$.

1.5 Let (X_n) be a sequence of independent random variables such that

$$\mathbb{P}[X_n = x] = \begin{cases} \frac{1}{n} & \text{if } x = n^2 \\ 1 - \frac{1}{n} & \text{if } x = 0. \end{cases}$$

Show that $\mathbb{P}[|X_n| > 0] \to 0$ and $\mathbb{E}[X_n] \to \infty$, as $n \to \infty$.

1.6 Let X be a normal random variable with mean μ and variance $\sigma^2 > 0$. By calculating $\mathbb{P}[Y \leq y]$ (or otherwise) show that $Y = \frac{X-\mu}{\sigma}$ is a normal random variable with mean 0 and variance 1.

1.7 For which values of $p \in (0, \infty)$ is $\int_1^\infty x^{-p}\,dx$ finite?

1.8 Which of the following sequences converge as $n \to \infty$? What do they converge too?

$$e^{-n} \qquad \sin\left(\frac{n\pi}{2}\right) \qquad \frac{\cos(n\pi)}{n} \qquad \sum_{i=1}^n 2^{-i} \qquad \sum_{i=1}^n \frac{1}{i}.$$

Give brief reasons for your answers.

1.9 Let (x_n) be a sequence of real numbers such that $\lim_{n\to\infty} x_n = 0$. Show that (x_n) has a subsequence (x_{n_r}) such that $\sum_{r=1}^\infty |x_{n_r}| < \infty$.

Chapter 2

Probability spaces and random variables

In this chapter we review probability theory, and develop some key tools for use in later chapters. We begin with a special focus on σ-fields. The role of a σ-field is to provide a way of controlling which information is visible (or, currently of interest) to us. As such, σ-fields will allow us to express the idea that, as time passes, we gain information.

2.1 Probability measures and σ-fields

Let Ω be a set. In probability theory, the symbol Ω is typically (and always, in this course) used to denote the *sample space*. Intuitively, we think of ourselves as conducting some random experiment, with an unknown outcome. The set Ω contains an $\omega \in \Omega$ for every possible outcome of the experiment. It is alternatively known as the *state space*.

Subsets of Ω correspond to collections of possible outcomes; such a subset is referred as an event. For instance, if we roll a dice we might take $\Omega = \{1, 2, 3, 4, 5, 6\}$ and the set $\{1, 3, 5\}$ is the event that our dice roll is an odd number.

Definition 2.1.1 Let $\mathcal{F}$ be a set of subsets of Ω. We say $\mathcal{F}$ is a σ-field if it satisfies the following properties:

1. $\emptyset \in \mathcal{F}$ and $\Omega \in \mathcal{F}$.
2. if $A \in \mathcal{F}$ then $\Omega \setminus A \in \mathcal{F}$.
3. if $A_1, A_2, \ldots \in \mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

The role of a σ-field is to choose which subsets of outcomes we are actually interested in. The power set $\mathcal{F} = \mathcal{P}(\Omega)$ is always a σ-field, and in this case every subset of Ω is an event. But $\mathcal{P}(\Omega)$ can be very big, and if our experiment is complicated, with many or even infinitely many possible outcomes, we might want to consider a smaller choice of $\mathcal{F}$ instead.

Sometimes we will need to deal with more than one σ-field at a time. A σ-field $\mathcal{G}$ such that $\mathcal{G} \subseteq \mathcal{F}$ is known as a *sub-σ-field* of $\mathcal{F}$.

We say that a subset $A \subseteq \Omega$ is *measurable*, or that it is an *event* (or *measurable event*), if $A \in \mathcal{F}$. To make to it clear which σ-field we mean to use in this definition, we sometimes write that an event is $\mathcal{F}$-measurable.

Example 2.1.2 Some examples of experiments and the σ-fields we might choose for them are the following:

- We toss a coin, which might result in heads H or tails T. We take $\Omega = \{H, T\}$ and $\mathcal{F} = \{\emptyset, \{H\}, \{T\}, \Omega\}$.
- We toss two coins, both of which might result in heads H or tails T. We take $\Omega = \{HH, TT, HT, TH\}$. However, we are only interested in the outcome that both coins are heads. We take $\mathcal{F} = \{\emptyset, \{HH\}, \Omega \setminus \{HH\}, \Omega\}$.

There are natural ways to choose a σ-field, even if we think of Ω as just an arbitrary set. For example, $\mathcal{F} = \{\Omega, \emptyset\}$ is a σ-field. If A is a subset of Ω, then $\mathcal{F} = \{\Omega, A, \Omega \setminus A, \emptyset\}$ is a σ-field (check it!).

Given Ω and $\mathcal{F}$, the final ingredient of a probability space is a measure $\mathbb{P}$, which tells us how likely the events in $\mathcal{F}$ are to occur.

Definition 2.1.3 A probability measure $\mathbb{P}$ is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ satisfying:

1. $\mathbb{P}[\Omega] = 1$.
2. If $A_1, A_2, \ldots \in \mathcal{F}$ are pair-wise disjoint (i.e. $A_i \cap A_j = \emptyset$ for all i, j such that $i \neq j$) then
$$\mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} \mathbb{P}[A_i].$$

The second of these conditions if often called σ-additivity. Note that we needed Definition 2.1.1 to make sense of Definion 2.1.3, because we needed something to tell us that $\mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i\right]$ was defined!

Definition 2.1.4 A probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ is a σ-field and $\mathbb{P}$ is a probability measure.

For example, to model a single fair coin toss we would take $\Omega = \{H, T\}$, $\mathcal{F} = \{\Omega, \{H\}, \{T\}, \emptyset\}$ and define $\mathbb{P}[H] = \mathbb{P}[T] = \frac{1}{2}$.

We commented above that often we want to choose $\mathcal{F}$ to be smaller than $\mathcal{P}(\Omega)$, but we have not yet shown how to choose a suitably small $\mathcal{F}$. Fortunately, there is a general way of doing so, for which we need the following technical lemma.

Lemma 2.1.5 *Let I be any set and for each $i \in I$ let $\mathcal{F}_i$ be a σ-field. Then*
$$\mathcal{F} = \bigcap_{i \in I} \mathcal{F}_i \tag{2.1}$$
is a σ-field

PROOF: We check the three conditions of Definition 2.1.1 for $\mathcal{F}$.

(1) Since each $\mathcal{F}_i$ is a σ-field, we have $\emptyset \in \mathcal{F}_i$. Hence $\emptyset \in \cap_i \mathcal{F}_i$. Similarly, $\Omega \in \mathcal{F}$.

(2) If $A \in \mathcal{F} = \cap_i \mathcal{F}_i$ then $A \in \mathcal{F}_i$ for each i. Since each $\mathcal{F}_i$ is a σ-field, $\Omega \setminus A \in \mathcal{F}_i$ for each i. Hence $\Omega \setminus A \in \cap_i \mathcal{F}_i$.

(3) If $A_j \in \mathcal{F}$ for all j, then $A_j \in \mathcal{F}_i$ for all i and j. Since each $\mathcal{F}_i$ is a σ-field, $\cup_j A_j \in \mathcal{F}_i$ for all i. Hence $\cup_j A_j \in \cap_i \mathcal{F}_i$. ■

Corollary 2.1.6 *In particular, if $\mathcal{F}_1$ and $\mathcal{F}_2$ are σ-fields, so is $\mathcal{F}_1 \cap \mathcal{F}_2$.*

Now, suppose that we have our Ω and we have a finite or countable collection of $E_1, E_2, \ldots \subseteq \Omega$, which we want to be events. Let $\mathscr{F}$ be the set of all σ-fields that contain $E_1, E_2, \ldots$. We enumerate $\mathscr{F}$ as $\mathscr{F} = \{\mathcal{F}_i\,;\, i \in I\}$, and apply Lemma 2.1.5. We thus obtain a σ-field $\mathcal{F}$, which contains all the events that we wanted.

The key point here is that $\mathcal{F}$ is the smallest σ-field that has $E_1, E_2, \ldots$ as events. To see why, note that by (2.1), $\mathcal{F}$ is contained inside any σ-field $\mathcal{F}'$ which has $E_1, E_2, \ldots$ as events.

Definition 2.1.7 Let $E_1, E_2, \ldots$ be subsets of Ω. We write $\sigma(E_1, E_2, \ldots,)$ for the smallest σ-field containing $E_1, E_2, \ldots$.

With Ω as any set, and $A \subseteq \Omega$, our example $\{\emptyset, A, \Omega \setminus A, \Omega\}$ is clearly $\sigma(A)$. In general, though, the point of Definition 2.1.7 is that we know useful σ-fields exist without having to construct them explicitly.

In the same style, if $\mathcal{F}_1, \mathcal{F}_2 \ldots$ are σ-fields then we write $\sigma(\mathcal{F}_1, \mathcal{F}_2, \ldots)$ for the smallest σ-algebra with respect to which all events in $\mathcal{F}_1, \mathcal{F}_2, \ldots$ are measurable.

From Definition 2.1.1 and 2.1.3 we can deduce all the 'usual' properties of probability. For example:

- If $A \in \mathcal{F}$ then $\Omega \setminus A \in \mathcal{F}$, and since $\Omega = A \cup (\Omega \setminus A)$ we have $\mathbb{P}[\Omega] = 1 = \mathbb{P}[A] + \mathbb{P}[\Omega \setminus A]$.
- If $A, B \in \mathcal{F}$ and $A \subseteq B$ then we can write $B = A \cup (B \setminus A)$, which gives us that $\mathbb{P}[B] = \mathbb{P}[B \setminus A] + P[A]$, which implies that $\mathbb{P}[A] \leq \mathbb{P}[B]$.

And so on. In this course we are concerned with applying probability theory rather than with relating its properties right back to the definition of a probability space; but you should realize that it is always possible to do so.

Definitions 2.1.1 and 2.1.3 both involve countable unions. It is convenient to be able to use countable intersections too, for which we need the following lemma.

Lemma 2.1.8 *Let $A_1, A_2, \ldots \in \mathcal{F}$, where $\mathcal{F}$ is a σ-field. Then $\bigcap_{i=1}^\infty A_i \in \mathcal{F}$.*

PROOF: We can write

$$\bigcap_{i=1}^\infty A_i = \bigcap_{i=1}^\infty \Omega \setminus (\Omega \setminus A_i) = \Omega \setminus \left(\bigcup_{i=1}^\infty \Omega \setminus A_i\right).$$

Since $\mathcal{F}$ is a σ-field, $\Omega \setminus A_i \in \mathcal{F}$ for all i. Hence also $\bigcup_{i=1}^\infty \Omega \setminus A_i \in \mathcal{F}$, which in turn means that $\Omega \setminus (\bigcup_{i=1}^\infty \Omega \setminus A_i) \in \mathcal{F}$. ■

In general, *uncountable* unions and intersections of measurable sets need not be measurable. The reasons why we only allow countable unions/intersections in probability are complicated and beyond the scope of this course. Loosely speaking, the bigger we make $\mathcal{F}$, the harder it is to make a probability measure $\mathbb{P}$, because we need to define $\mathbb{P}[A]$ for all $A \in \mathcal{F}$ in a way that satisfies Definition 2.1.3. Allowing uncountable set operations would (in natural situations) result in $\mathcal{F}$ being so large that it would be *impossible* to find a suitable $\mathbb{P}$.

From now on, the symbols Ω, $\mathcal{F}$ and $\mathbb{P}$ always denote the three elements of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

2.2 Random variables

Our probability space gives us a label $\omega \in \Omega$ for every possible outcome. Sometimes it is more convenient to think about a property of ω, rather than about ω itself. For this, we use a random variable, $X : \Omega \to \mathbb{R}$. For each outcome $\omega \in \Omega$, the value of $X(\omega)$ is a property of the outcome.

For example, let $\Omega = \{1, 2, 3, 4, 5, 6\}$ and $\mathcal{F} = \mathcal{P}(\Omega)$. We might be interested in the property

$$X(\omega) = \begin{cases} 0 & \text{if } \omega \text{ is odd,} \\ 1 & \text{if } \omega \text{ is even.} \end{cases}$$

We write

$$X^{-1}(A) = \{\omega \in \Omega\,;\, X(\omega) \in A\},$$

for $A \subseteq \mathbb{R}$, which is called the *pre-image* of A under X. In words, $X^{-1}(A)$ is the set of outcomes ω for which the property $X(\omega)$ falls inside the set A. In our example above $X^{-1}(\{0\}) = \{1, 3, 5\}$, $X^{-1}(\{1\}) = \{2, 4, 6\}$ and $X^{-1}(\{0, 1\}) = \{1, 2, 3, 4, 5, 6\}$.

It is common to write $X^{-1}(a)$ in place of $X^{-1}(\{a\})$, because it makes easier reading. Similarly, for an interval $(a, b) \subseteq \mathbb{R}$ we write $X^{-1}(a, b)$ in place of $X^{-1}\big((a, b)\big)$.

Definition 2.2.1 Let $\mathcal{G}$ be a σ-field. A function $X : \Omega \to \mathbb{R}$ is said to be $\mathcal{G}$-measurable if

$$\text{for all subintervals } I \subseteq \mathbb{R}, \quad \text{we have } X^{-1}(I) \in \mathcal{G}.$$

If it is clear which σ-field $\mathcal{G}$ we mean to use, which might simply say that X is *measurable.* We will often shorten this to writing simply $X \in m\mathcal{G}$.

For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we say that $X : \Omega \to \mathbb{R}$ is a *random variable* if X is $\mathcal{F}$-measurable. The relationship to the usual notation for probability is that $\mathbb{P}[X \in A]$ means $\mathbb{P}[X^{-1}(A)]$, so as e.g.

$$\mathbb{P}\left[a < X < b\right] = \mathbb{P}\left[X^{-1}(a, b)\right] = \mathbb{P}[\omega \in \Omega\,;\, X(\omega) \in (a, b)]$$
$$\mathbb{P}[X = a] = \mathbb{P}[X^{-1}(a)] = \mathbb{P}[\omega \in \Omega\,;\, X(\omega) = a].$$

We usually prefer writing $\mathbb{P}[X = a]$ and $\mathbb{P}[a < X < b]$ because we find them more intuitive; we like to think of X as an object that takes a random value.

For example, suppose we toss a coin twice, with $\Omega = \{HH, HT, TH, TT\}$ as in Example 2.1.2. If we take our σ-field to be $\mathcal{F} = \mathcal{P}(\Omega)$, the set of all subsets of Ω, then any function $X : \Omega \to \mathbb{R}$ is $\mathcal{F}$-measurable. However, suppose we choose instead

$$\mathcal{G} = \{\Omega, \{HT, TH, TT\}, \{HH\}, \emptyset\}$$

(as we did in Example 2.1.2). Then if we look at function

$$X(\omega) = \text{ the total number of tails which occurred}$$

we have $X^{-1}([0, 1]) = \{HH, HT, TH\} \notin \mathcal{G}$, so X is not $\mathcal{G}$-measurable. The intuition here is that σ-field $\mathcal{G}$ 'isn't big enough' for X, because $\mathcal{G}$ only contains information about whether we threw two heads (or not).

σ-fields and information

It will be very important for us to understand the connection between σ-fields and information. When we talk about the 'information' contained in a σ-field $\mathcal{G}$, we mean the following.

Suppose that an outcome ω of our experiment has occurred, but suppose that we don't know which $\omega \in \Omega$ it was. Each event $G \in \mathcal{G}$ represents a piece of information. This piece of information is whether or not $\omega \in G$ i.e. whether or not the event G has occurred. If this 'information' allows us to deduce the exact value of $X(\omega)$, and if we can do this for any $\omega \in \Omega$, then X is $\mathcal{G}$-measurable.

Going back to our example above, of two coin tosses, the information contained in $\mathcal{G}$ is whether (or not) we threw two heads. Recall that $X \in \{0, 1, 2\}$ was the number of tails thrown. Knowing just this information provided by $\mathcal{G}$ doesn't allow us to deduce X – for example if all we know is that we didn't throw two heads, we can't work out exactly how many tails we threw.

The interaction between random variables and σ-fields can be summarised as follows:

$$\begin{array}{rcl}
\boxed{\text{event}} & \leftrightarrow & \boxed{\text{a piece of information: did the event occur}} \\
\boxed{\sigma\text{-field } \mathcal{G}} & \leftrightarrow & \boxed{\text{which information we care about}} \\
\boxed{X \text{ is } \mathcal{G}\text{-measurable}} & \leftrightarrow & \boxed{X \text{ depends only on information that we care about}}
\end{array}$$

Rigorously, if we want to check that X is $\mathcal{G}$-measurable, we have to check that $X^{-1}(I) \in \mathcal{G}$ for every subinterval of $I \subseteq \mathbb{R}$. This can be tedious, especially if X takes many different values. Fortunately, we will shortly see that, in practice, there is rarely any need to do so. What is important for us is to understand the role played by a σ-field.

σ-fields and pre-images

Suppose that we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X : \Omega \to \mathbb{R}$. We want to check that X is measurable with respect to some smaller σ-field $\mathcal{G}$.

If X is a discrete random variable, we can use the following lemma.

Lemma 2.2.2 *Let $\mathcal{G}$ be a σ-field on Ω. Let $X : \Omega \to \mathbb{R}$, and suppose X takes a finite or countable set of values $\{x_1, x_2, \ldots\}$. Then:*

$$X \textit{ is measurable with respect to } \mathcal{G} \quad \Leftrightarrow \quad \textit{for all } j,\ \{X = x_j\} \in \mathcal{G}.$$

Proof: Let us first prove the $\Rightarrow$ implication. So, assume the left hand side: $X \in m\mathcal{G}$. Since $\{X = x_j\} = X^{-1}(x_j)$ and $\{x_j\} = [x_j, x_j]$ is a subinterval of $\mathbb{R}$, by Definition 2.2.1 we have that $\{X = x_j\} \in \mathcal{G}$. Since this holds for any $j = 1, \ldots, n$, we have shown the right hand side.

Next, we will prove the $\Leftarrow$ implication. So, we now assume the right hand side: $\{X = x_j\} \in \mathcal{G}$ for all j. Let I be any subinterval of $\mathbb{R}$, and define $J = \{j\,;\, x_j \in I\}$. Then,

$$X^{-1}(I) = \{\omega \in \Omega\,;\, X(\omega) \in I\} = \bigcup_{j \in J} \{\omega\,;\, X(w) = x_j\} = \bigcup_{j \in J} \{X = x_j\}.$$

Since $\{X = x_j\} \in \mathcal{G}$, the definition of a σ-field tell us that also $X^{-1}(I) \in \mathcal{G}$. ∎

For example, take $\Omega = \{1, 2, 3, 4, 5, 6\}$, which we think of as rolling a dice, and take $\mathcal{F} = \mathbb{P}(\Omega)$. Consider

$$\begin{aligned} \mathcal{G}_1 &= \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}, \\ \mathcal{G}_2 &= \{\emptyset, \{1, 2, 3\}, \{4, 5, 6\}, \Omega\}, \end{aligned}$$

Here, $\mathcal{G}_1$ contains the information of whether the roll is even or odd, and $\mathcal{G}_2$ contains the information of whether the roll is ≤ 3 or > 3. It's easy to check that $\mathcal{G}_1$ and $\mathcal{G}_2$ are both σ-fields.

Let us define

$$X_1(\omega) = \begin{cases} 0 & \text{if } \omega \text{ is odd,} \\ 1 & \text{if } \omega \text{ is even,} \end{cases}$$

$$X_2(\omega) = \begin{cases} 0 & \text{if } \omega \leq 3, \\ 1 & \text{if } \omega > 3. \end{cases}$$

That is, X_1 tests if the roll is even, and X_2 tests if the roll is less than or equal to three. Based on our intuition about information we should expect that X_1 is measurable with respect to $\mathcal{G}_1$ but not $\mathcal{G}_2$, and that X_2 is measurable with respect to $\mathcal{G}_2$ but not $\mathcal{G}_1$.

We can justify our intuition rigorously using Lemma 2.2.2. We have $X_1^{-1}(0) = \{1, 3, 5\}$, $X_1^{-1}(1) = \{2, 4, 6\}$, so $X_1 \in m\mathcal{G}_1$ but $X_1 \notin m\mathcal{G}_2$. Similarly, we have $X_2^{-1}(0) = \{1, 2, 3\}$ and $X_2^{-1}(1) = \{4, 5, 6\}$, so $X_2 \notin m\mathcal{G}_1$ and $X_2 \in m\mathcal{G}_2$.

Let us extend this example by introducing a third σ-field,

$$\mathcal{G}_3 = \sigma\big(\{1, 3\}, \{2\}, \{4\}, \{5\}, \{6\}\big).$$

The σ-field $\mathcal{G}_3$ is, by Definition 2.1.7, the smallest σ-field containing the events $\{1, 3\}, \{2\}, \{4\}, \{5\}$ and $\{6\}$. It contains the information of which $\omega \in \Omega$ we threw *except* that it can't tell the difference between a 1 and a 3. If we tried to write $\mathcal{G}_3$ out in full we would discover that it had 32 elements (and probably make some mistakes!) so instead we just use Definition 2.1.7.

To check if $X_1 \in m\mathcal{G}_3$, we need to check if $\mathcal{G}_3$ contains $\{1,3,5\}$ and $\{2,4,6\}$. We can write

$$\{1,3,5\} = \{1,3\} \cup \{5\}, \qquad \{2,4,6\} = \{2\} \cup \{4\} \cup \{6\}$$

which shows that $\{1,3,5\}, \{2,4,6\} \in \mathcal{G}_3$ because, in both cases, the right hand sides are made up of sets that we already know are in $\mathcal{G}_3$, plus countable set operations. Hence, X_1 is $\mathcal{G}_3$ measurable. You can check for yourself that X_2 is also $\mathcal{G}_3$ measurable.

Remark 2.2.3 ($\oslash$) For continuous random variables, there is no equivalent of Lemma 2.2.2. More sophisticated tools from measure theory are needed – see MAS350/61022.

σ-fields generated by random variables

We can think of random variables as containing information, because their values tell us something about the result of the experiment. We can express this idea formally: there is a natural σ-field associated to each function $X : \Omega \to \mathbb{R}$.

Definition 2.2.4 The σ-field generated by X, denoted $\sigma(X)$, is

$$\sigma(X) = \sigma(X^{-1}(I)\,;\, I \text{ is a subinterval of } \mathbb{R}).$$

In words, $\sigma(X)$ is the σ-field generated by the sets $X^{-1}(I)$ for intervals I. The intuition is that $\sigma(X)$ is the smallest σ-field of events on which the random behaviour of X depends.

For example, consider throwing a fair die. Let $\Omega = \{1, 2, 3, 4, 5, 6\}$, let $\mathcal{F} = \mathcal{P}(\Omega)$ and let

$$X(\omega) = \begin{cases} 1 & \text{if } \omega \text{ is odd} \\ 2 & \text{if } \omega \text{ is even.} \end{cases}$$

Then $X(\omega) \in \{1, 2\}$, with pre-images $X^{-1}(1) = \{1, 3, 5\}$ and $X^{-1}(2) = \{2, 4, 6\}$. The smallest σ-field that contains both of these subsets is

$$\sigma(X) = \Big\{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\Big\}.$$

The information contained in this σ-field is whether ω is even or odd, which is precisely the same information given by the value of $X(\omega)$.

In general, if X takes lots of different values, $\sigma(X)$ could be very big and we would have no hope of writing it out explicitly. Here's another example: suppose that

$$Y(\omega) = \begin{cases} 1 & \text{if } \omega = 1, \\ 2 & \text{if } \omega = 2, \\ 3 & \text{if } \omega \geq 3. \end{cases}$$

Then $Y(\omega) \in \{1, 2, 3\}$ with pre-images $Y^{-1}(1) = \{1\}$, $Y^{-1}(2) = \{2\}$ and $Y^{-1}(3) = \{3, 4, 5, 6\}$. The smallest σ-field containing these three subsets is

$$\sigma(Y) = \Big\{\emptyset, \{1\}, \{2\}, \{3, 4, 5, 6\}, \{1, 2\}, \{1, 3, 4, 5, 6\}, \{2, 3, 4, 5, 6\}, \Omega\Big\}.$$

The information contained in this σ-field is whether ω is equal to 1, 2 or some number ≥ 3. Again, this is precisely the same information as is contained in the value of $X(\omega)$.

It's natural that X should be measurable with respect to the σ-field that contains precisely the information on which X depends. Formally:

Lemma 2.2.5 *Let $X : \Omega \to \mathbb{R}$. Then X is $\sigma(X)$-measurable.*

PROOF: Let I be a subinterval of $\mathbb{R}$. Then, by definition of $\sigma(X)$, we have that $X^{-1}(I) \in \sigma(X)$ for all I. ∎

If we have a finite or countable set of random variables $X_1, X_2, \ldots$ we define $\sigma(X_1, X_2, \ldots)$ to be $\sigma(X_1^{-1}(I), X_2^{-1}(I), \ldots\,;\, I \text{ is a subinterval of } \mathbb{R})$. The intuition is the same: $\sigma(X_1, X_2, \ldots)$ corresponds to the information jointly contained in $X_1, X_2, \ldots$.

Combining random variables

Given a collection of random variables, it is useful to be able to construct other random variables from them. To do so we have the following proposition. Since we will eventually deal with more than one σ-field at once, it is useful to express this idea for a sub-σ-field $\mathcal{G} \subseteq \mathcal{F}$.

Proposition 2.2.6 *Let $\alpha \in \mathbb{R}$ and let $X, Y, X_1, X_2, \ldots$ be $\mathcal{G}$-measurable functions from $\Omega \to \mathbb{R}$. Then*

$$\alpha, \quad \alpha X, \quad X+Y, \quad XY, \quad 1/X, \tag{2.2}$$

are all $\mathcal{G}$-measurable[1]. *Further, if X_∞ given by*

$$X_\infty(\omega) = \lim_{n\to\infty} X_n(\omega)$$

exists for all ω, then X_∞ is $\mathcal{G}$-measurable.

Essentially, every natural way of combining random variables together leads to other random variables. Proposition 2.2.6 can usually be used to show this. We won't prove Proposition 2.2.6 in this course, see MAS31002/61022.

For example, if X is a random variable then so is $\frac{X^2+X}{2}$. For a more difficult example, suppose that X is a random variable and let $Y = e^X$, which means that $Y(\omega) = \lim_{n\to\infty} \sum_{i=0}^n \frac{X(\omega)^i}{i!}$. Recall that we know from analysis that this limit exists since $e^x = \lim_{n\to\infty} \sum_{i=0}^n \frac{x^i}{i!}$ exists for all $x \in \mathbb{R}$. Each of the partial sums

$$Y_n = \sum_{i=0}^n \frac{X^i}{i!} = 1 + X + \frac{X^2}{2} + \ldots + \frac{X^n}{n!}$$

is a random variable (we could use (2.2) repeatedly to show this) and, since the limit exists, $Y(\omega) = \lim_{n\to\infty} Y_n(\omega)$ is measurable.

In general, if X is a random variable and $g : \mathbb{R} \to \mathbb{R}$ is any 'sensible' function then $g(X)$ is also a random variable. This includes polynomials, powers, all trig functions, all monotone functions, all continuous and piecewise continuous functions, integrals/derivatives, etc etc.

Independence

We can express the concept of independence, which you already know about for random variables, in terms of σ-fields. Recall that two events $E_1, E_2 \in \mathcal{F}$ are said to be independent if $\mathbb{P}[E_1 \cap E_2] = \mathbb{P}[E_1]\mathbb{P}[E_2]$. Using σ-fields, we have a consistent way of defining independence, for both random variables and events.

Definition 2.2.7 Sub-σ-fields $\mathcal{G}_1, \mathcal{G}_2$ of $\mathcal{F}$ are said to be independent if $\mathbb{P}(G_1 \cap G_2) = \mathbb{P}(G_1)\mathbb{P}(G_2)$ for all $G_1 \in \mathcal{G}_1$ and $G_2 \in \mathcal{G}_2$.

Events E_1 and E_2 are independent if $\sigma(E_1)$ and $\sigma(E_2)$ are independent.

Random variables X_1 and X_2 are independent if $\sigma(X_1)$ and $\sigma(X_2)$ are independent.

It can be checked that, for events and random variables, this definition is equivalent to the definitions you will have seen in earlier courses. The same principle, of using the associated σ-fields, applies to defining what it means for e.g. a random variable and an event to be independent. We won't check this claim as part of our course, see MAS31002/61022.

[1] In the case of $1/X$, we require that $\mathbb{P}[X = 0] = 0$.

2.3 Infinite Ω

So far we focused on finite sample spaces of the form $\Omega = \{x_1, x_2, \ldots x_n\}$. In such a case we would normally take $\mathcal{F} = \mathcal{P}(\Omega)$, which is also a finite set. Since $\mathcal{F}$ contains every subset of Ω, any σ-field on Ω is a sub-σ-field of $\mathcal{F}$. We have seen how it is possible to construct other σ-fields on Ω too.

In this case we can define a probability measure on Ω by choosing a finite sequence $a_1, a_2, \ldots, a_n$ such that each $a_i \in [0,1]$ and $\sum_1^n a_i = 1$. We set $\mathbb{P}[x_i] = a_i$. This naturally extends to defining $\mathbb{P}[A]$ for any subset $A \subseteq \Omega$, by setting

$$\mathbb{P}[A] = \sum_{\{i\,;\,x_i \in A\}} \mathbb{P}[x_i] = \sum_{\{i\,;\,x_i \in A\}} a_i. \tag{2.3}$$

It is hopefully obvious (and tedious to check) that, with this definition, $\mathbb{P}$ is a probability measure. Consequently $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

All our experiments with (finitely many tosses/throws of) dice and coins fit into this category of examples. In fact, if our experiment has countably many outcomes, say, $\Omega = \{x_1, x_2, \ldots\}$ we can still work in much the same way, and the sum in (2.3) will become an infinite series that sums to 1.

However, the theory of stochastic processes, as well as most sophisticated examples of stochastic models, require an *uncountable* sample space. In such cases, we can't use (2.3), because there is no such thing as an uncountable sum.

An example with uncountable Ω

We now flex our muscles a bit, and look at an example where Ω is uncountable. We toss a coin infinitely many times, then $\Omega = \{H, T\}^{\mathbb{N}}$, meaning that we write an outcome as a sequence $\omega = \omega_1, \omega_2, \ldots$ where $\omega_i \in \{H, T\}$. The set Ω is uncountable.

We define the random variables $X_n(\omega) = \omega_n$, so as X_n represents the result (H or T) of the n^{th} throw. We take

$$\mathcal{F} = \sigma(X_1, X_2, \ldots)$$

i.e. $\mathcal{F}$ is the smallest σ-field with respect to which all the X_n are random variables. Then

$$\begin{aligned}
\sigma(X_1) &= \{\emptyset, \{H\star\star\star\ldots\}, \{T\star\star\star\ldots\}, \Omega\} \\
\sigma(X_1, X_2) &= \sigma\Big(\{HH\star\star\ldots\}, \{TH\star\star\ldots\}, \{HT\star\star\ldots\}, \{TT\star\star\ldots\}\Big) \\
&= \Big\{\emptyset, \{HH\star\star\ldots\}, \{TH\star\star\ldots\}, \{HT\star\star\ldots\}, \{TT\star\star\ldots\}, \\
&\qquad \{H\star\star\star\ldots\}, \{T\star\star\star\ldots\}, \{\star H\star\star\ldots\}, \{\star T\star\star\ldots\}, \left\{\begin{array}{c} HH\star\star\ldots \\ TT\star\star\ldots \end{array}\right\}, \left\{\begin{array}{c} HT\star\star\ldots \\ TH\star\star\ldots \end{array}\right\}, \\
&\qquad \{HH\star\star\ldots\}^c, \{TH\star\star\ldots\}^c, \{HT\star\star\ldots\}^c, \{TT\star\star\ldots\}^c, \Omega\Big\},
\end{aligned}$$

where $\star$ means that it can take on either H or T, so $\{H\star\star\star\ldots\} = \{\omega : \omega_1 = H\}$.

With the information available to us in $\sigma(X_1, X_2)$, we can distinguish between ω's where the first or second outcomes are different. But if two ω's have the same first and second outcomes, they fall into exactly the same subset(s) of $\sigma(X_1, X_2)$. Consequently, if a random variable depends on anything more than the first and second outcomes, it will not be $\sigma(X_1, X_2)$ measurable.

It is not immediately clear if we can define a probability measure on $\mathcal{F}$! Since Ω is uncountable, we cannot use the idea of (2.3) and define $\mathbb{P}$ in terms of $\mathbb{P}[\omega]$ for each individual $\omega \in \Omega$. Equation (2.3) simply would not make sense; there is no such thing as an uncountable sum.

To define a probability measure in this case requires a significant amount of machinery from measure theory. It is outside of the scope of this course. For our purposes, whenever we need to use an infinite Ω you will be *given* a probability measure and some of its helpful properties. For example, in this case there exists a probability measure $\mathbb{P} : \mathcal{F} \to [0, 1]$ such that

- The X_n are independent random variables.
- $\mathbb{P}[X_n = H] = \mathbb{P}[X_n = T] = \frac{1}{2}$ for all $n \in \mathbb{N}$.

From this, you can work with $\mathbb{P}$ without having to know how $\mathbb{P}$ was constructed. You don't even need to know exactly which subsets of Ω are in $\mathcal{F}$, because Proposition 2.2.6 gives you access to plenty of random variables.

Remark 2.3.1 ($\oslash$) In this case it turns out that $\mathcal{F}$ is much smaller than $\mathcal{P}(\Omega)$. In fact, if we tried to take $\mathcal{F} = \mathcal{P}(\Omega)$, we would (after some significant effort) discover that there is no probability measure $\mathbb{P} : \mathcal{P}(\Omega) \to [0, 1]$ (i.e. satisfying Definition 2.1.3) in which the coin tosses are independent. This is irritating, and surprising, and we just have to live with it.

Almost surely

In the example from Section 2.3 we used $\Omega = \{H, T\}^{\mathbb{N}}$, which is the set of all sequences made up of Hs and Ts. Our probability measure was independent, fair, coin tosses and we used the random variable X_n for the n^{th} toss.

Let's examine this example a bit. First let us note that, for any individual sequence $\omega_1, \omega_2, \dots$ of heads and tails, by independence

$$\mathbb{P}[X_1 = \omega_1, X_2 = \omega_2, \dots] = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \dots = 0.$$

So every element of Ω has probability zero. This is not a problem – if we take enough elements of Ω together then we do get non-zero probabilities, for example

$$\mathbb{P}[X_1 = H] = \mathbb{P}\left[\omega \in \Omega \text{ such that } \omega_1 = H\right] = \frac{1}{2}$$

which is not surprising.

The probability that we never throw a head is

$$\mathbb{P}[\text{for all } n, X_n = T] = \frac{1}{2} \cdot \frac{1}{2} \dots = 0$$

which means that the probability that we eventually throw a head is

$$\mathbb{P}[\text{for some } n, X_n = H] = 1 - \mathbb{P}[\text{for all } n, X_n = T] = 1.$$

So, the event $\{\text{for some } n, X_n = H\}$ has probability 1, but is *not equal to* the whole sample space Ω. To handle this situation we have a piece of terminology.

Definition 2.3.2 If the event E has $\mathbb{P}[E] = 1$, then we say E occurs almost surely.

So, we would say that 'almost surely, our coin will eventually throw a head'. We might say that '$Y \leq 1$' almost surely, to mean that $\mathbb{P}[Y \leq 1] = 1$. This piece of terminology will be used very frequently from now on. We might sometimes say that an event 'almost always' happens, with the same meaning.

For another example, suppose that we define q_n^H and q_n^T to be the proportion of heads and, respectively, tails in the first n coin tosses $X_1, X_2, \dots, X_n$. Formally, this means that

$$q_n^H = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i = H\} \quad \text{and} \quad q_n^T = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i = T\}.$$

Here $\mathbb{1}\{X_i = H\}$ is equal to 1 is $X_i = H$, and equal to zero otherwise; similarly for $\mathbb{1}\{X_i = T\}$. We will think a bit more about random variables of this type in the next section. Of course $q_n^H + q_n^T = 1$.

The random variables $\mathbb{1}\{X_i = H\}$ are i.i.d. with $\mathbb{E}[\mathbb{1}\{X_i = H\}] = \frac{1}{2}$, hence by Theorem 1.1.1 we have $\mathbb{P}[q_n^H \to \frac{1}{2} \text{ as } n \to \infty] = 1$, and by the same argument we have also $\mathbb{P}[q_n^T \to \frac{1}{2} \text{ as } n \to \infty] = 1$. In words, this means that in the long run half our tosses will be tails and half will be heads (which makes sense - our coin is fair). That is, the event

$$E = \left\{\lim_{n\to\infty} q_n^H = \frac{1}{2} \text{ and } \lim_{n\to\infty} q_n^T = \frac{1}{2}\right\}$$

occurs almost surely.

There are many many examples of sequences (e.g. $HHTHHTHHT\dots$) that don't have $q_n^T \to \frac{1}{2}$ and $q_n^H \to \frac{1}{2}$. We might think of the event E as being only a 'small' subset of Ω, but it has probability one.

2.4 Expectation

There is only one part of the 'usual machinery' for probability that we haven't yet discussed, namely expectation.

Recall that the expectation of a discrete random variable X that takes the values $\{x_i : i \in \mathbb{N}\}$ is given by

$$\mathbb{E}[X] = \sum_{x_i} x_i \mathbb{P}[X = x_i]. \tag{2.4}$$

For a continuous random variables, the expectation uses an integral against the probability density function,

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x\, f_X(x)\, dx. \tag{2.5}$$

Recall also that it is possible for limits (i.e. infinite sums) and integrals to be infinite, or not exist at all.

We are now conscious of the general definition of a random variable X, as an $\mathcal{F}$-measurable function from Ω to $\mathbb{R}$. There are many random variables that are neither discrete nor continuous, and for such cases (2.4) and (2.5) are not valid; we need a more general approach.

With *Lebesgue integration*, the expectation $\mathbb{E}$ can be defined using a single definition that works for both discrete and continuous (and other more exotic) random variables. This definition relies heavily on analysis and is well beyond the scope of this course. Instead, Lebesgue integration is covered in MAS31002/61022.

For purposes of this course, what you should know is: $\mathbb{E}[X]$ is defined for all X such that either

1. $X \geq 0$, in which case it is possible that $\mathbb{E}[X] = \infty$,
2. general X for which $\mathbb{E}[|X|] < \infty$.

The point here is that we are prepared to allow ourselves to write $\mathbb{E}[X] = \infty$ (e.g. when the sum or integral in (2.4) or (2.5) tends to ∞) *provided* that $X \geq 0$. We are not prepared to allow expectations to equal $-\infty$, because we have to avoid nonsensical '$\infty - \infty$' situations.

It's worth knowing that if $X \geq 0$ and $\mathbb{P}[X = \infty] > 0$, then $\mathbb{E}[X] = \infty$. In words, the slightest chance of X being infinite will outweigh all of the finite possibilities and make $\mathbb{E}[X]$ infinite.

You may still use (2.4) and (2.5), in the discrete/continuous cases. You may also assume that all the 'standard' properties of $\mathbb{E}$ hold:

Proposition 2.4.1 *For random variables X, Y:*

(Linearity) *If $a, b \in \mathbb{R}$ then* $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.
(Independence) *If X and Y are independent then* $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.
(Absolute values) $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.
(Monotonicity) *If $X \leq Y$ then* $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

You should become familiar with any of the properties that you are not already used to using. The proofs of these properties are part of the formal construction of $\mathbb{E}$ and are not part of our course.

Indicator functions

One important type of random variable is an indicator function. Let $A \in \mathcal{F}$, then the indicator function of A is the function

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A. \end{cases}$$

The indicator function is used to tell if an event occurred (in which case it is 1) or did not occur (in which case it is 0). It is useful to remember that

$$\mathbb{P}[A] = \mathbb{E}[\mathbb{1}_A].$$

We will sometimes not put the A as a subscript and write e.g. $\mathbb{1}\{X < 0\}$ for the indicator function of the event that $X < 0$.

As usual, let $\mathcal{G}$ denote a sub σ-field of $\mathcal{F}$.

Lemma 2.4.2 *Let $A \in \mathcal{G}$. Then the function $\mathbb{1}_A$ is $\mathcal{G}$-measurable.*

Proof: Let us write $Y = \mathbb{1}_A$. Note that Y is a discrete random variable, which can take the two values 0 and 1. We have $Y^{-1}(1) = \{Y = 1\} = A$ and $Y^{-1}(0) = \{Y = 0\} = \Omega \setminus A$. By Proposition 2.2.2, Y is $\mathcal{G}$ measurable. ∎

Indicator functions allow us to *condition*, meaning that we can break up a random variable into two or more cases. For example, given any random variable X we can write

$$X = X\mathbb{1}_{\{X \geq 1\}} + X\mathbb{1}_{\{X<1\}}. \tag{2.6}$$

Precisely one of the two terms on the right hand side is non-zero. If $X \geq 1$ then the first term takes the value X and the second is zero; if $X < 1$ then the second term is equal to X and the first term is zero.

We can use (2.6) to prove a useful inequality. Putting $|X|$ in place of X, and then taking $\mathbb{E}$ we obtain

$$\begin{aligned} \mathbb{E}[|X|] &= \mathbb{E}[|X|\mathbb{1}_{\{|X|\geq 1\}}] + \mathbb{E}[|X|\mathbb{1}_{\{|X|<1\}}] \\ &\leq \mathbb{E}[X^2\mathbb{1}_{\{|X|\geq 1\}}] + 1 \\ &\leq \mathbb{E}[X^2] + 1. \end{aligned} \tag{2.7}$$

Here, to deduce the second line, the key point is we can only use the inequality $|x| \leq x^2$ if $x \geq 1$. Hence, for the first term we can use that $|X|\mathbb{1}_{\{|X|\geq 1\}} \leq X^2\mathbb{1}_{\{|X|\geq 1\}}$. For the second term, we use that $|X|\mathbb{1}_{\{|X|<1\}} \leq 1$. In both cases, we also need the monotonicity of $\mathbb{E}$.

L^p spaces

It will often be important for to us know whether a random variable X has finite mean and variance. Some random variables do not, see exercise **2.8** (or MAS2010 or MAS31002) for example. Random variables with finite mean and variances are easier to work with than those which don't, and many of the results in this course require these conditions.

We use some notation:

Definition 2.4.3 Let $p \in [1, \infty)$. We say that $X \in L^p$ if $\mathbb{E}[|X|^p] < \infty$.

In this course, we will only be interested in the cases $p = 1$ and $p = 2$. These cases have the following set of useful properties:

1. By definition, L^1 is the set of random variables for which $\mathbb{E}[|X|]$ is finite.
2. L^2 is the set of random variables with finite variance. This comes from the fact that $\mathrm{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ (more strictly, it needs some integration theory from MAS31002/61022).
3. From (2.7), if $X \in L^2$ then also $X \in L^1$.

Often, to check if $X \in L^p$ we must calculate $\mathbb{E}[|X|^p]$. A special case where it is automatic is the following.

Definition 2.4.4 We say that a random variable X is bounded if there exists (deterministic) $c \in \mathbb{R}$ such that $|X| \le c$.

If X is bounded, then using monotonicity we have $\mathbb{E}[|X|^p] \le \mathbb{E}[c^p] = c^p < \infty$, which means that $X \in L^p$, for all p.

2.5 Exercises on Chapter 2

On probability spaces

2.1 Consider the experiment of throwing two dice, then recording the uppermost faces of both dice. Write down a suitable sample space Ω and suggest an appropriate σ-field $\mathcal{F}$.

2.2 Let $\Omega = \{1, 2, 3\}$. Let

$$\mathcal{F} = \{\emptyset, \{1\}, \{2, 3\}, \{1, 2, 3\}\},$$
$$\mathcal{F}' = \{\emptyset, \{2\}, \{1, 3\}, \{1, 2, 3\}\}.$$

(a) Show that $\mathcal{F}$ and $\mathcal{F}'$ are both σ-fields.

(b) Show that $\mathcal{F} \cup \mathcal{F}'$ is not a σ-field, but that $\mathcal{F} \cap \mathcal{F}'$ is a σ-field.

2.3 Let $\Omega = \{H, T\}^{\mathbb{N}}$ be the probability space from Section 2.3, corresponding to an infinite sequence of independent fair coin tosses $(X_n)_{n=1}^{\infty}$.

(a) Fix $m \in \mathbb{N}$. Show that the probability that that random sequence $X_1, X_2, \ldots,$ contains precisely m heads is zero.

(b) Deduce that, almost surely, the sequence $X_1, X_2, \ldots$ contains infinitely many heads and infinitely many tails.

2.4 Let $\Omega = \{1, 2, 3, 4, 5, 6\}$, corresponding to one roll of a die. In each of the following cases describe, in words, the information contained within the given σ-field.

(a) $\mathcal{F}_1 = \{\emptyset, \{1, 2, 3, 4, 5\}, \{6\}, \Omega\}$.

(b) $\mathcal{F}_2 = \sigma(\{1\}, \{2\}, \{3\}, \{4, 5, 6\})$.

(c) $\mathcal{F}_3 = \{\emptyset, \{1, 2\}, \{3, 4\}, \{5, 6\}, \{1, 2, 3, 4\}, \{3, 4, 5, 6\}, \{1, 2, 5, 6\}, \Omega\}$.

On random variables

2.5 Let $\Omega = \{1, 2, 3, 4, 5\}$ and set

$$\mathcal{G}_1 = \sigma(\{1, 5\}, \{2, 4\}, \{3\})$$
$$\mathcal{G}_2 = \{\emptyset, \{1, 2\}, \{3, 4, 5\}, \Omega\}.$$

(a) Define $X_1 : \Omega \to \mathbb{R}$ by $X_1(\omega) = (\omega - 3)^2$. Show that $X_1 \in m\mathcal{G}_1$ and $X_1 \notin m\mathcal{G}_2$.

(b) Give an example of a function $X_2 : \Omega \to \mathbb{R}$ such that $X_2 \in m\mathcal{G}_2$ and $X_2 \notin m\mathcal{G}_1$.

2.6 Let $\Omega = \{HH, HT, TH, TT\}$, representing two coin tosses. Define X to be the total number of heads shown. Write down all the events in $\sigma(X)$.

2.7 Let X be a random variable. Explain why $\frac{X}{X^2+1}$ and $\sin(X)$ are also random variables.

2.8 Let X be a random variable with the probability density function $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} 2x^{-3} & \text{if } x \in [1, \infty), \\ 0 & \text{otherwise.} \end{cases}$$

Show that $X \in L^1$ but $X \notin L^2$.

2.9 Let $1 \leq p \leq q < \infty$.

(a) Generalize (2.7) to show that $\mathbb{E}[|X|^p] \leq \mathbb{E}[|X|^q] + 1$.

(b) Deduce that if $X \in L^q$ then $X \in L^p$.

2.10 Let $a > 0$ and let X be a random variable such that $X \geq 0$. Show that $\mathbb{P}[X \geq a] \leq \frac{1}{a}\mathbb{E}[X]$.

Challenge questions

2.11 Show that if $\mathbb{P}[X \geq 0] = 1$ and $\mathbb{E}[X] = 0$ then $\mathbb{P}[X = 0] = 1$.

Chapter 3

Conditional expectation and martingales

We will introduce conditional expectation, which provides us with a way to estimate random quantities based on only partial information. We will also introduce martingales, which are the mathematical way to capture the concept of a fair game.

3.1 Conditional expectation

Suppose X and Z are random variables that take on only finitely many values $\{x_1, \ldots, x_m\}$ and $\{z_1, \ldots, z_n\}$, respectively. In earlier courses, 'conditional expectation' was defined as follows:

$$\begin{aligned}
\mathbb{P}[X = x_i \,|\, Z = z_j] &= \mathbb{P}[X = x_i, Z = z_j]/\mathbb{P}[Z = z_j] \\
\mathbb{E}[X \,|\, Z = z_j] &= \sum_i x_i \mathbb{P}[X = x_i \,|\, Z = z_j] \\
Y = \mathbb{E}[X \,|\, Z] \text{ where:} &\qquad \text{if } Z(\omega) = z_j, \text{ then } Y(\omega) = \mathbb{E}[X \,|\, Z = z_j] \qquad (3.1)
\end{aligned}$$

You might also have seen a second definition, using probability density functions, for continuous random variables. These definitions are problematic, for several reasons, chiefly (1) its not immediately clear how the two definitions interact and (2) we don't want to be restricted to handling only discrete or only continuous random variables.

In this section, we define the conditional expectation of random variables using σ-fields. In this setting we are able to give a unified definition which is valid for general random variables. The definition is originally due to Kolmogorov (in 1933), and is sometimes referred to as Kolmogorov's conditional expectation. It is one of the most important concepts in modern probability theory.

Conditional expectation is a mathematical tool with the following function. We have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X : \Omega \to \mathbb{R}$. However, $\mathcal{F}$ is large and we want to work with a sub-σ-algebra $\mathcal{G}$, instead. As a result, we want to have a random variable Y such that

1. Y is $\mathcal{G}$-measurable
2. Y is 'the best' way to approximate X with a $\mathcal{G}$-measurable random variable

The second statement on this wish-list does not fully make sense; there are many different ways in which we could compare X to a potential Y.

Why might we want to do this? Imagine we are conducting an experiment in which we gradually gain information about the result X. This corresponds to gradually seeing a larger and larger $\mathcal{G}$, with access to more and more information. At all times we want to keep a prediction of what the future looks like, based on the currently available information. This prediction is Y.

It turns out there is *only one* natural way in which to realize our wish-list (which is convenient, and somewhat surprising). It is the following:

Theorem 3.1.1 (Conditional Expectation) *Let X be an L^1 random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$. Then there exists a random variable $Y \in L^1$ such that*

1. *Y is $\mathcal{G}$-measurable,*
2. *for every $G \in \mathcal{G}$, we have $\mathbb{E}[Y\mathbb{1}_G] = \mathbb{E}[X\mathbb{1}_G]$.*

Moreover, if $Y' \in L^1$ is a second random variable satisfying these conditions, $\mathbb{P}[Y = Y'] = 1$.

The first and second statements here correspond respectively to the items on our wish-list.

Definition 3.1.2 We refer to Y as (a version of) the *conditional expectation* of X given $\mathcal{G}$. and we write

$$Y = \mathbb{E}[X \mid \mathcal{G}].$$

Since any two such Y are almost surely equal so we sometimes refer to Y simply as *the* conditional expectation of X. This is a slight abuse of notation, but it is commonplace and harmless.

Proof of Theorem 3.1.1 is beyond the scope of this course. Loosely speaking, there is an abstract recipe which constructs $\mathbb{E}[X|\mathcal{G}]$. It begins with the random variable X, and then averages out over all the information that is not accessible to $\mathcal{G}$, leaving only as much randomness as $\mathcal{G}$ can support, resulting in $\mathbb{E}[X|\mathcal{G}]$. In this sense the map $X \mapsto \mathbb{E}[X|\mathcal{G}]$ simplifies (i.e. reduces the amount of randomness in) X in a very particular way, to make it $\mathcal{G}$ measurable.

It is important to remember that $\mathbb{E}[X|\mathcal{G}]$ is (in general) a random variable. It is also important to remember that the two objects

$$\mathbb{E}[X|\mathcal{G}] \quad \text{and} \quad \mathbb{E}[X|Z = z]$$

are quite different. They are both useful. We will explore the connection between them in Section 3.1. Before doing so, let us look at a basic example.

Let X_1, X_2 be independent random variables such that $\mathbb{P}[X_i = -1] = \mathbb{P}[X_i = 1] = \frac{1}{2}$. Set $\mathcal{F} = \sigma(X_1, X_2)$. We will show that

$$\mathbb{E}[X_1 + X_2|\sigma(X_1)] = X_1. \tag{3.2}$$

To do so, we should check that X_1 satisfies the two conditions in Theorem 3.1.1, with

$$\begin{aligned} X &= X_1 + X_2 \\ Y &= X_1 \\ \mathcal{G} &= \sigma(X_1). \end{aligned}$$

The first condition is immediate, since by Lemma 2.2.5 X_1 is $\sigma(X_1)$-measurable i.e. $Y \in m\mathcal{G}$. To see the second condition, let $G \in \sigma(X_1)$. Then $\mathbb{1}_G \in \sigma(X_1)$ by Lemma 2.4.2 and $X_2 \in \sigma(X_2)$, and these σ-fields are independent, so $\mathbb{1}_G$ and X_2 are independent. Hence

$$\begin{aligned}
\mathbb{E}[(X_1 + X_2)\mathbb{1}_G] &= \mathbb{E}[X_1\mathbb{1}_G] + \mathbb{E}[1_G X_2] \\
&= \mathbb{E}[X_1\mathbb{1}_G] + \mathbb{E}[1_G]\mathbb{E}[X_2] \\
&= \mathbb{E}[X_1\mathbb{1}_G] + \mathbb{P}[G].0 \\
&= \mathbb{E}[X_1\mathbb{1}_G].
\end{aligned}$$

This equation says precisely that $\mathbb{E}[X\mathbb{1}_G] = \mathbb{E}[Y\mathbb{1}_G]$. We have now checked both conditions, so by Theorem 3.1.1 we have $\mathbb{E}[X|\mathcal{G}] = Y$, meaning that $\mathbb{E}[X_1 + X_2|\sigma(X_1)] = X_1$, which proves our claim in (3.2).

The intuition for this, which is plainly visible in our calculation, is that X_2 is independent of $\sigma(X_1)$ so, thinking of conditional expectation as an operation which averages out all randomness in $X = X_1 + X_2$ that is not $\mathcal{G} = \sigma(X_1)$ measurable, we would average out X_2 completely i.e. $\mathbb{E}[X_2] = 0$.

We could equally think of X_1 as being our best guess for $X_1 + X_2$, given only information in $\sigma(X_1)$, since $\mathbb{E}[X_2] = 0$. In general, guessing $\mathbb{E}[X|\mathcal{G}]$ is not so easy!

Relationship to the naive definition (⊘)

Conditional expectation extends the 'naive' definition of (3.1). Naturally, the 'new' conditional expectation is much more general (and, moreover, it is what we require later in the course), but we should still take the time to relate it to the naive definition.

Remark 3.1.3 This subsection is marked with a (⊘), meaning that it is non-examinable. This is so as you can forget the old definition and remember the new one!

To see the connection, we focus on the case where X, Z are random variables with finite sets of values $\{x_1, \dots, x_n\}$, $\{z_1, \dots, z_m\}$. Let Y be the naive version of conditional expectation defined in (3.1). That is,

$$Y(\omega) = \sum_j \mathbb{1}_{\{Z(\omega)=z_j\}} \mathbb{E}[X|Z = z_j].$$

We can use Theorem 3.1.1 to check that, in fact, Y is a version of $\mathbb{E}[X|\sigma(Z)]$. We want to check that Y satisfies the two properties listed in Theorem 3.1.1.

- Since Z only takes finitely many values $\{z_1, \dots, z_m\}$, from the above equation we have that Y only takes finitely many values. These values are $\{y_1, \dots, y_m\}$ where $y_j = \mathbb{E}[X|Z = z_j]$. We note

  $$\begin{aligned} Y^{-1}(y_j) &= \{\omega \in \Omega\,;\, Y(\omega) = \mathbb{E}[X|Z = z_j]\} \\ &= \{\omega \in \Omega\,;\, Z(\omega) = z_j\} \\ &= Z^{-1}(z_j) \in \sigma(Z). \end{aligned}$$

  This is sufficient (although we will omit the details) to show that Y is $\sigma(Z)$-measurable.

- We can calculate

  $$\begin{aligned} \mathbb{E}[Y\mathbb{1}\{Z = z_j\}] &= y_j \mathbb{E}[\mathbb{1}\{Z = z_j\}] \\ &= y_j \mathbb{P}[Z = z_j] \\ &= \sum_i x_i \mathbb{P}[X = x_i | Z = z_j] \mathbb{P}[Z_j = z_j] \\ &= \sum_i x_i \mathbb{P}[X = x_i \text{ and } Z = z_j] \\ &= \sum_{i,j} x_i \mathbb{1}_{\{Z=z_j\}} \mathbb{P}[X = x_i \text{ and } Z = z_j] \\ &= \mathbb{E}[X\mathbb{1}_{\{Z=z_j\}}]. \end{aligned}$$

  Properly, to check that Y satisfies the second property in Theorem 3.1.1, we need to check $\mathbb{E}[Y\mathbb{1}_G] = \mathbb{E}[X\mathbb{1}_G]$ for a general $G \in \sigma(Z)$ and not just $G = \{Z = z_j\}$. However, for reasons beyond the scope of this course, in this case (thanks to the fact that Z is finite) its enough to consider only G of the form $\{Z = z_j\}$.

Therefore, we have $Y = \mathbb{E}[X|\sigma(Z)]$ almost surely. In this course we favour writing $\mathbb{E}[X|\sigma(Z)]$ instead of $\mathbb{E}[X|Z]$, to make it clear that we are looking at conditional expectation with respect to a σ-field.

3.2 Properties of conditional expectation

In all but the easiest cases, calculating conditional expectations explicitly from Theorem 3.1.1 is not feasible. Instead, we are able to work with them via a set of useful properties, provided by the following proposition.

Proposition 3.2.1 *Let $\mathcal{G},\mathcal{H}$ be sub-σ-fields of $\mathcal{F}$ and $X,Y,Z\in L^1$. Then, almost surely,*

(Linearity) $\mathbb{E}[a_1X_1+a_2X_2\,|\,\mathcal{G}]=a_1\mathbb{E}[X_1\,|\,\mathcal{G}]+a_2\mathbb{E}[X_2\,|\,\mathcal{G}]$.
(Absolute values) $|\mathbb{E}[X\,|\,\mathcal{G}]|\leq\mathbb{E}[|X|\,|\,\mathcal{G}]$.
(Montonicity) *If $X\leq Y$, then $\mathbb{E}[X|\mathcal{G}]\leq\mathbb{E}[Y|\mathcal{G}]$.*
(Constants) *If $a\in\mathbb{R}$ (deterministic) then $\mathbb{E}[a\,|\,\mathcal{G}]=a$.*

(Measurability) *If X is $\mathcal{G}$-measurable, then $\mathbb{E}[X\,|\,\mathcal{G}]=X$.*
(Independence) *If X is independent of $\mathcal{G}$ then $\mathbb{E}[X\,|\,\mathcal{G}]=\mathbb{E}[X]$.*

(Taking out what is known) *If Z is $\mathcal{G}$ measurable, then $\mathbb{E}[ZX\,|\,\mathcal{G}]=Z\mathbb{E}[X\,|\,\mathcal{G}]$.*
(Tower) *If $\mathcal{H}\subset\mathcal{G}$ then $\mathbb{E}[\mathbb{E}[X\,|\,\mathcal{G}]\,|\,\mathcal{H}]=\mathbb{E}[X\,|\,\mathcal{H}]$.*
(Taking $\mathbb{E}$) *It holds that $\mathbb{E}[\mathbb{E}[X\,|\,\mathcal{G}]]=\mathbb{E}[X]$.*
(No information) *It holds that $\mathbb{E}[X\,|\,\{\emptyset,\Omega\}]=\mathbb{E}[X]$,*

Proof of these properties is beyond the scope of our course. Note that the first five properties above are common properties of both $\mathbb{E}[\cdot]$ and $\mathbb{E}[\cdot\,|\,\mathcal{G}]$.

We'll use these properties extensively, for the whole of the remainder of the course. They are not on the formula sheet – you should remember them and become familiar with applying them.

Remark 3.2.2 (⊘) Although we have not proved the properties in Proposition 3.2.1, they are intuitive properties for conditional expectation to have.

For example, in the taking out what is known property, we can think of Z as already being simple enough to be $\mathcal{G}$ measurable, so we'd expect that taking conditional expectation with respect to $\mathcal{G}$ doesn't need to affect it.

In the independence property, we can think of $\mathcal{G}$ as giving us no information about the value X is taking, so our best guess at the value of X has to be simply $\mathbb{E}[X]$.

In the tower property for $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]$, we start with X, simplify it to be $\mathcal{G}$ measurable and simplify it to be $\mathcal{H}$ measurable. But since $\mathcal{H}\subseteq\mathcal{G}$, we might as well have just simplified X enough to be $\mathcal{H}$ measurable in a single step, which would be $\mathbb{E}[X|\mathcal{H}]$.

Etc. It is a useful exercise for you to try and think of 'intuitive' arguments for the other properties too, so as you can easily remember them.

Conditional expectation as an estimator

The conditional expectation $Y = \mathbb{E}[X \mid \mathcal{G}]$ is the 'best least-squares estimator' of X, based on the information available in $\mathcal{G}$. We can state this rigorously and use our toolkit from Proposition 3.2.1 prove it. It demonstrates another way in which Y is 'the best' $\mathcal{G}$-measurable approximation to X, and provides our first example of using the properties of $\mathbb{E}[X \mid \mathcal{G}]$.

Lemma 3.2.3 *Let $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$. Let X be an $\mathcal{F}$-measurable random variable and let $Y = \mathbb{E}[X|\mathcal{G}]$. Suppose that Y' is a $\mathcal{G}$-measurable, random variable. Then*

$$\mathbb{E}[(X-Y)^2] \le \mathbb{E}[(X-Y')^2].$$

PROOF: We note that

$$\begin{aligned}\mathbb{E}[(X-Y')^2] &= \mathbb{E}[(X-Y+Y-Y')^2] \\ &= \mathbb{E}[(X-Y)^2] + 2\mathbb{E}[(X-Y)(Y-Y')] + \mathbb{E}[(Y-Y')^2]. \end{aligned} \tag{3.3}$$

In the middle term above, we can write

$$\begin{aligned}\mathbb{E}[(X-Y)(Y-Y')] &= \mathbb{E}[\mathbb{E}[(X-Y)(Y-Y')|\mathcal{G}]] \\ &= \mathbb{E}[(Y-Y')\mathbb{E}[X-Y|\mathcal{G}]] \\ &= \mathbb{E}[(Y-Y')(\mathbb{E}[X|\mathcal{G}]-Y)] \\ &= \mathbb{E}[(Y-Y')(0)] \\ &= 0.\end{aligned}$$

Here, in the first step we used the 'taking $\mathbb{E}$' property, in the second step we used Proposition 2.2.6 to tell us that $Y - Y'$ is $\mathcal{G}$-measurable, followed by the 'taking out what is known' rule. In the final step we used the linearity and measurability properties. Since $\mathbb{E}[X|\mathcal{G}] = Y$ almost surely, we obtain that $\mathbb{E}[(X-Y)(Y-Y')] = 0$. Hence, since $\mathbb{E}[(Y-Y')^2] \ge 0$, from (3.3) we obtain $\mathbb{E}[(X-Y')^2] \ge \mathbb{E}[(X-Y)^2]$. ∎

3.3 Martingales

In this section we introduce martingales, which are the mathematical representation of a 'fair game'. As usual, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We refer to a sequence of random variables $(S_n)_{n=0}^\infty$ as a stochastic process. In this section of the course we only deal with discrete time stochastic processes.

We have previously discussed the idea of gradually learning more and more information about the outcome of some experiment, through seeing the information visible from gradually larger σ-fields. We formalize this concept as follows.

Definition 3.3.1 A sequence of σ-fields $(\mathcal{F}_n)_{n=0}^\infty$ is known as a filtration if $\mathcal{F}_0 \subseteq \mathcal{F}_1 \ldots \subseteq \mathcal{F}$.

Definition 3.3.2 We say that a stochastic process $X = (X_n)$ is adapted to the filtration $(\mathcal{F}_n)$ if, for all n, X_n is $\mathcal{F}_n$ measurable.

We should think of the filtration $\mathcal{F}_n$ as telling us which information we have access too at time $n = 1, 2, \ldots$. Thus, an adapted process is a process whose (random) value we know at all times $n \in \mathbb{N}$.

We are now ready to give the definition of a martingale.

Definition 3.3.3 A process $M = (M_n)_{n=0}^\infty$ is a martingale if

1. if (M_n) is adapted,
2. $M_n \in L^1$ for all n,
3. $\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n] = M_n$ almost surely, for all n.

We say that M is a submartingale if, instead of 3, we have $\mathbb{E}[M_n \,|\, \mathcal{F}_{n-1}] \geq M_{n-1}$ almost surely. We say that M is a supermartingale if, instead of 3, we have $\mathbb{E}[M_n \,|\, \mathcal{F}_{n-1}] \leq M_{n-1}$ almost surely.

Remark 3.3.4 The second condition in Definition 3.3.3 is needed for the third to make sense.

Remark 3.3.5 (M_n) is a martingale iff it is both a submartingale and a supermartingale.

A martingale is the mathematical idealization of a fair game. It is best to understand what we mean by this through an example.

Let (X_n) be a sequence of i.i.d. random variables such that

$$\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = -1] = \frac{1}{2}.$$

Define $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ and $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Then $(\mathcal{F}_n)$ is a filtration. Define

$$S_n = \sum_{i=1}^n X_i$$

(and $S_0 = 0$). We can think of S_n as a game in the following way. At each time $n = 1, 2, \ldots$ we toss a coin. We win if the n^{th} round if the coin is heads, and lose if it is tails. Each time we win we score 1, each time we lose we score -1. Thus, S_n is our score after n rounds. The process S_n is often called a simple random walk.

We claim that S_n is a martingale. To see this, we check the three properties in the definition. (1) Since $X_1, X_2, \ldots, X_n \in m\sigma(X_1, \ldots, X_n)$ we have that $S_n \in \mathcal{F}_n$ for all $n \in \mathbb{N}$. (2) Since $|S_n| \le n$ for all $n \in \mathbb{N}$, $\mathbb{E}[|S_n|] \le n$ for all n, so $S_n \in L^1$ for all n. (3) We have

$$\begin{aligned}\mathbb{E}\left[S_{n+1} \,|\, \mathcal{F}_n\right] &= \mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n] + \mathbb{E}[S_n \,|\, \mathcal{F}_n] \\ &= \mathbb{E}[X_{n+1}] + S_n \\ &= S_n.\end{aligned}$$

Here, in the first line we used the linearity of conditional expectation. To deduce the second line we used the relationship between independence and conditional expectation (for the first term) and the measurability rule (for the second term). To deduce the final line we used that $\mathbb{E}[X_{n+1}] = (1)\frac{1}{2} + (-1)\frac{1}{2} = 0$.

At time n we have seen the result of rounds $1, 2, \ldots, n$, so the information we currently have access to is given by $\mathcal{F}_n$. This means that at time n we know $S_1, \ldots, S_n$. But we don't know S_{n+1}, because S_{n+1} is not $\mathcal{F}_n$-measurable. However, using our current information we can make our best guess at what S_{n+1} will be, which naturally is $\mathbb{E}[S_{n+1}|\mathcal{F}_n]$. Since the game is fair, in the future, on average we do not expect to win more than we lose, that is $\mathbb{E}[S_{n+1}|\mathcal{F}_n] = S_n$.

In this course we will see many examples of martingales, and we will gradually build up an intuition for how to recognize a martingale. There is, however, one easy sufficient (but not necessary) condition under which we can recognize that a stochastic process is not a martingale.

Lemma 3.3.6 *Let $(\mathcal{F}_n)$ be a filtration and suppose that (M_n) is a martingale. Then*

$$\mathbb{E}[M_n] = \mathbb{E}[M_0]$$

for all $n \in \mathbb{N}$.

PROOF: We have $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n$. Taking expectations and using the 'taking $\mathbb{E}$' property from Proposition 3.2.1, we have $\mathbb{E}[M_{n+1}] = \mathbb{E}[M_n]$. The result follows by a trivial induction. ■

Suppose, now, that (X_n) is an i.i.d. sequence of random variables such that $\mathbb{P}[X_i = 2] = \mathbb{P}[X_i = -1] = \frac{1}{2}$. Note that $\mathbb{E}[X_n] > 0$. Define S_n and $\mathcal{F}_n$ as before. Now, $\mathbb{E}[S_n] = \sum_1^n \mathbb{E}[X_n]$, which is not constant, so S_n is not a martingale. However, as before, S_n is $\mathcal{F}_n$-measurable, and $|S_n| \le 2n$ so $S_n \in L^1$, essentially as before. We have

$$\begin{aligned}\mathbb{E}[S_{n+1}|\mathcal{F}_n] &= \mathbb{E}[X_{n+1}|\mathcal{F}_n] + \mathbb{E}[S_n|\mathcal{F}_n] \\ &= \mathbb{E}[X_{n+1}] + S_n \\ &\ge S_n.\end{aligned}$$

Hence S_n is a submartingale.

In general, if (M_n) is a submartingale, then by definition $\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n] \ge M_n$, so taking expectations gives us $\mathbb{E}[M_{n+1}] \ge \mathbb{E}[M_n]$. For supermartingales we get $\mathbb{E}[M_{n+1}] \le \mathbb{E}[M_n]$. In words: submartingales, on average, increase, whereas supermartingales, on average, decrease. The use of super- and sub- is counter intuitive in this respect.

Remark 3.3.7 Sometimes we will want to make it clear which filtration is being used in the definition of a martingale. To do so we might say that (M_n) is an $\mathcal{F}_n$-martingale', or that (M_n) is a martingale with respect to $\mathcal{F}_n$'. We use the same notation for super/sub-martingales.

Our definition of a filtration and a martingale both make sense if we look at only a finite set of times $n = 1, \dots, N$. We sometimes also use the terms filtration and martingale in this situation.

We end this section with two important general examples of martingales. You should check the conditions yourself, as exercise **3.3**.

Example 3.3.8 Let (X_n) be a sequence of i.i.d. random variables such that $\mathbb{E}[X_n] = 1$ for all n, and there exists $c \in \mathbb{R}$ such that $|X_n| \le c$ for all n. Define $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. Then

$$M_n = \prod_{i=1}^{n} X_n$$

is a martingale.

Example 3.3.9 Let $Z \in L^1$ be a random variable and let $(\mathcal{F}_n)$ be a filtration. Then

$$M_n = \mathbb{E}[Z|\mathcal{F}_n]$$

is a martingale.

3.4 Exercises on Chapter 3

On conditional expectation and martingales

3.1 Let (X_n) be a sequence of independent identically distributed random variables, such that $\mathbb{P}[X_i = 1] = \frac{1}{2}$ and $\mathbb{P}[X_i = -1] = \frac{1}{2}$. Let

$$S_n = \sum_{i=1}^{n} X_i.$$

Use the properties of conditional expectation to find $\mathbb{E}[S_2 \,|\, \sigma(X_1)]$ and $\mathbb{E}[S_2^2 \,|\, \sigma(X_1)]$ in terms of X_1, X_2 and their expected values.

3.2 Let (X_n) be a sequence of independent random variables such that $\mathbb{P}[X_n = 2] = \frac{1}{3}$ and $\mathbb{P}[X_n = -1] = \frac{2}{3}$. Set $\mathcal{F}_n = \sigma(X_i\,;\, i \leq n)$. Show that $S_n = \sum_{i=1}^{n} X_i$ is an $\mathcal{F}_n$ martingale.

3.3 Check that Examples 3.3.8 and 3.3.9 are martingales.

3.4 Let (M_t) be a stochastic process that is both and submartingale and a supermartingale. Show that (M_t) is a martingale.

3.5 (a) Let (M_n) be an $\mathcal{F}_n$ martingale. Show that, for all $0 \leq n \leq m$, $\mathbb{E}[M_m \,|\, \mathcal{F}_n] = M_n$.

(b) Guess and state (without proof) the analogous result to (a) for submartingales.

3.6 Let (M_n) be a $\mathcal{F}_n$ martingale and suppose $M_n \in L^2$ for all n. Show that

$$\mathbb{E}[M_{n+1}^2|\mathcal{F}_n] = M_n^2 + \mathbb{E}[(M_{n+1} - M_n)^2|\mathcal{F}_n] \tag{3.4}$$

and deduce that (M_n^2) is a submartingale.

3.7 Let $X_0, X_1, \ldots$ be a sequence of L^1 random variables. Let $\mathcal{F}_n$ be their generated filtration and suppose that $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = aX_n + bX_{n-1}$ for all $n \in \mathbb{N}$, where $a, b > 0$ and $a + b = 1$.

Find a value of $\alpha \in \mathbb{R}$ (in terms of a, b) for which $S_n = \alpha X_n + X_{n-1}$ is an $\mathcal{F}_n$ martingale.

Challenge questions

3.8 In the setting of **3.1**, show that $\mathbb{E}[X_1 \,|\, \sigma(S_n)] = \frac{S_n}{n}$.

Chapter 4

Stochastic processes

In this chapter we introduce stochastic processes, with a selection of examples that are commonly used as building blocks in stochastic modelling. We show that these stochastic processes are closely connected to martingales.

Definition 4.0.1 A stochastic process (in discrete time) is a sequence $(X_n)_{n=0}^{\infty}$ of random variables. We think of n as 'time'.

For example, a sequence of i.i.d. random variables is a stochastic process. A martingale is a stochastic process. A Markov chain (from MAS2003, for those who took it) is a stochastic process. And so on.

For any stochastic process (X_n) the *natural* or *generated filtration* of (X_n) is the filtration given by

$$\mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n).$$

Therefore, a random variable is $\mathcal{F}_m$ measurable if it depends only on the behaviour of our stochastic process up until time m.

From now on we adopt the convention (which is standard in the field of stochastic processes) that whenever we don't specify a filtration explicitly we mean to use the generated filtration.

4.1 Random walks

Random walks are stochastic processes that 'walk around' in space. We think of a particle that moves between vertices of $\mathbb{Z}$. At each step of time, the particle chooses at random to either move up or down, for example from x to $x+1$ or $x-1$.

Simple symmetric random walk

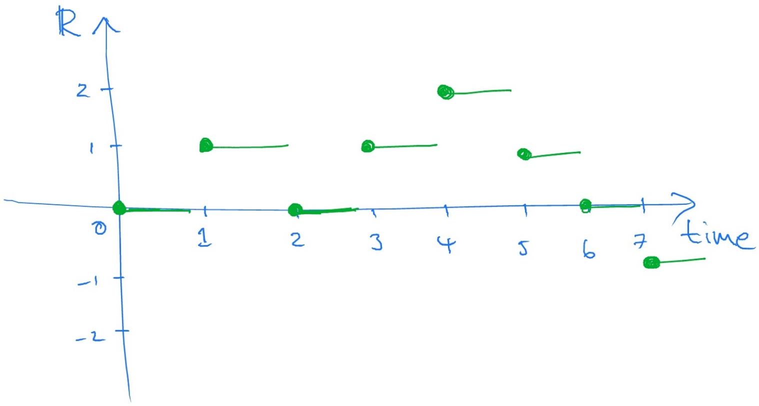
Let $(X_i)_{i=1}^{\infty}$ be a sequence of i.i.d. random variables where

$$\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = -1] = \frac{1}{2}. \tag{4.1}$$

The simple symmetric random walk is the stochastic process

$$S_n = \sum_{i=1}^{n} X_i.$$

By convention, this means that $S_0 = 0$. The word 'simple' refers to the fact that the walk moves by precisely 1 unit of space in each step of time i.e. $|X_i| = 1$. A *sample path* of S_n, which means a sample of the sequence $(S_0, S_1, S_2, \ldots)$, might look like:



Note that when time is discrete $t = 0, 1, 2, \ldots$ it is standard to draw the location of the random walk (and other stochastic processes) as constant in between integer time points.

Because of (4.1), the random walk is equally likely to move upwards or downwards. This case is known as the 'symmetric' random walk because, if $S_0 = 0$, the two stochastic processes S_n and $-S_n$ have the same distribution.

We have already seen (in Section 3.3) that S_n is a martingale, with respect to its generated filtration

$$\mathcal{F}_n = \sigma(X_1, \ldots, X_n) = \sigma(S_1, \ldots, S_n).$$

It should seem very natural that (S_n) is a martingale – going upwards as much as downwards is 'fair'.

Simple asymmetric random walk

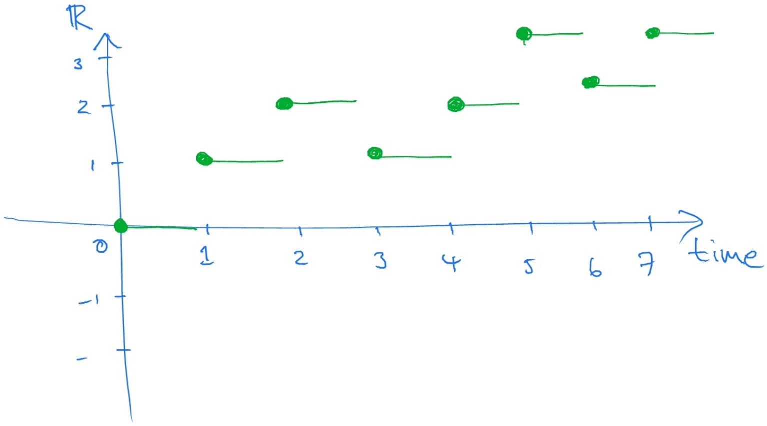
Let $(X_i)_{i=1}^{\infty}$ be a sequence of i.i.d. random variables. Let $p + q = 1$ with $p, q \in [0, 1]$, $p \neq q$ and suppose that

$$\mathbb{P}[X_i = 1] = p, \quad \mathbb{P}[X_i = -1] = q.$$

The asymmetric random walk is the stochastic process

$$S_n = \sum_{i=1}^{n} X_i.$$

The key difference to the symmetric random walk is that here we have $p \neq q$ (the symmetric random walk has $p = q = \frac{1}{2}$). The asymmetric random walk is more likely to step upwards than downwards if $p > q$, and vice versa if $q < p$. The technical term for this behaviour is *drift*. A sample path for the case $p > q$ might look like:



This is 'unfair', because of the drift upwards, so we should suspect that the asymmetric random walk is not a martingale. In fact,

$$\mathbb{E}[S_n] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} (p-q) = n(p-q), \tag{4.2}$$

whereas $\mathbb{E}[S_0] = 0$. Thus, Lemma 3.3.6 confirms that S_n is not a martingale. However, the process

$$M_n = S_n - n(p-q) \tag{4.3}$$

is a martingale. The key is that the term $n(p-q)$ *compensates* for the drift and 'restores fairness'.

We'll now prove that (M_n) is a martingale. Since $X_i \in m\mathcal{F}_n$ for all $i \le n$, by Proposition 2.2.6 we have $S_n - n(p-q) \in m\mathcal{F}_n$. Since $|X_i| \le 1$ we have

$$|S_n - n(p-q)| \le |S_n| + n|p-q| \le n + n|p-q|$$

and hence M_n is bounded, so $M_n \in L^1$. Lastly,

$$\begin{aligned}
\mathbb{E}[S_{n+1} - (n+1)(p-q) \,|\, \mathcal{F}_n] &= \mathbb{E}[S_{n+1} \,|\, \mathcal{F}_n] - (n+1)(p-q) \\
&= \mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n] + \mathbb{E}[S_n \,|\, \mathcal{F}_n] - (n+1)(p-q) \\
&= \mathbb{E}[X_{n+1}] + S_n - (n+1)(p-q) \\
&= (p-q) + S_n - (n+1)(p-q) \\
&= S_n - n(p-q).
\end{aligned}$$

Therefore $\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n] = M_n$, and (M_n) is a martingale.

4.2 Urn processes

Urn processes are 'balls in bags' processes. In the simplest kind of urn process, which we look at in this section, we have just a single urn (i.e. bag) that contains balls of two different colours.

At time 0, an urn contains 1 black ball and 1 red ball. Then, for each $n = 1, 2, \ldots$, we generate the state of the urn at time n by doing the following:

1. Draw a ball from the urn, look at its colour, and return this ball to the urn.

2. Add a new ball of the same colour as the drawn ball.

So, at time n (which means: after the n^{th} iteration of the above steps is completed) there are $n+2$ balls in the urn. This process is known as the *Pólya* urn. Pólya was a Hungarian mathematician who made contributions across a wide spectrum mathematics, in the first half of the 20^{th} century.

Let B_n be the number of red balls in the urn at time n, and note that $B_0 = 1$. Set $(\mathcal{F}_n)$ to be the filtration generated by (B_n).

Our first step is to note that B_n itself is *not* a martingale. The reason is that over time we will put more and more red balls into the urn, so the number of red balls drifts upwards over time. Formally, we can note that

$$\begin{aligned}
\mathbb{E}[B_{n+1}\,|\,\mathcal{F}_n] &= \mathbb{E}[B_{n+1}\mathbb{1}_{\{(n+1)^{th}\text{ draw is red}\}}\,|\,\mathcal{F}_n] + \mathbb{E}[B_{n+1}\mathbb{1}_{\{(n+1)^{th}\text{ draw is black}\}}\,|\,\mathcal{F}_n] \\
&= \mathbb{E}[(B_n+1)\mathbb{1}_{\{(n+1)^{th}\text{ draw is red}\}}\,|\,\mathcal{F}_n] + \mathbb{E}[B_n\mathbb{1}_{\{(n+1)^{th}\text{ draw is black}\}}\,|\,\mathcal{F}_n] \\
&= (B_n+1)\mathbb{E}[\mathbb{1}_{\{(n+1)^{th}\text{ draw is red}\}}\,|\,\mathcal{F}_n] + B_n\mathbb{E}[\mathbb{1}_{\{(n+1)^{th}\text{ draw is black}\}}\,|\,\mathcal{F}_n] \\
&= (B_n+1)\frac{B_n}{n+2} + B_n\left(1 - \frac{B_n}{n+2}\right) \\
&= \frac{B_n(n+3)}{n+2} > B_n. \qquad (4.4)
\end{aligned}$$

We do have $B_n \in m\mathcal{F}_n$ and since $1 \le B_n \le n+2$ we also have $B_n \in L^1$, so B_n is a submartingale, but due to (4.4) B_n is not a martingale.

However, a closely related quantity is a martingale. Let

$$M_n = \frac{B_n}{n+2}.$$

Then M_n is the proportion of balls in the urn that are red, at time n. Note that $M_n \in [0,1]$. We can think of the extra factor $n+2$, which increases over time, as an attempt to cancel out the upwards drift of B_n. We now have:

$$\begin{aligned}
\mathbb{E}[M_{n+1}\,|\,\mathcal{F}_n] &= \mathbb{E}\left[M_{n+1}\mathbb{1}_{\{(n+1)^{th}\text{ draw is red}\}}\,\middle|\,\mathcal{F}_n\right] + \mathbb{E}\left[M_{n+1}\mathbb{1}_{\{(n+1)^{th}\text{ draw is black}\}}\,\middle|\,\mathcal{F}_n\right] \\
&= \mathbb{E}\left[\frac{B_n+1}{n+3}\mathbb{1}_{\{(n+1)^{th}\text{ draw is red}\}}\,\middle|\,\mathcal{F}_n\right] + \mathbb{E}\left[\frac{B_n}{n+3}\mathbb{1}_{\{(n+1)^{th}\text{ draw is black}\}}\,\middle|\,\mathcal{F}_n\right] \\
&= \frac{B_n+1}{n+3}\mathbb{E}\left[\mathbb{1}_{\{(n+1)^{th}\text{ draw is red}\}}\,\middle|\,\mathcal{F}_n\right] + \frac{B_n}{n+3}\mathbb{E}\left[\mathbb{1}_{\{(n+1)^{th}\text{ draw is black}\}}\,\middle|\,\mathcal{F}_n\right] \\
&= \frac{B_n+1}{n+3}\frac{B_n}{n+2} + \frac{B_n}{n+3}\left(1 - \frac{B_n}{n+2}\right) \\
&= \frac{B_n^2+B_n}{(n+2)(n+3)} + \frac{(n+2)B_n - B_n^2}{(n+2)(n+3)}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{(n+3)B_n}{(n+2)(n+3)} \\
&= \frac{B_n}{n+2} \\
&= M_n.
\end{aligned}
$$

We have $M_n \in m\mathcal{F}_n$ and since $M_n \in [0,1]$ we have that $M_n \in L^1$. Hence (M_n) is a martingale.

We can think of $M_n = \frac{B_n}{n+2}$ as a compensation mechanism; B_n tends to increase, and we compensate for this increase by dividing by $n+2$.

Remark 4.2.1 The calculation of $\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n]$ is written out in full as a second example of the method. In fact, we could simply have divided the equality in (4.4) by $n+3$, and obtained $\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n] = M_n$.

On fairness

It is clear that the symmetric random walk is fair; at all times it is equally likely to move up as down. The asymmetric random walk is not fair, due to its drift (4.2), but once we compensate for drift in (4.3) we do still obtain a martingale.

Then urn process requires more careful thought. For example, we might wonder:

> *Suppose that the first draw is red. Then, at time $n = 1$ we have two red balls and one black ball. So, the chance of drawing a red ball is now $\frac{2}{3}$. How is this fair?!*

To answer this question, let us make a number of points. Firstly, let us remind ourselves that the quantity which is a martingale is M_n, the *proportion* of red balls in the urn.

Secondly, suppose that the first draw is indeed red. So, at $n = 1$ we have 2 red and 1 black, giving a proportion of $\frac{2}{3}$ red and $\frac{1}{3}$ black. The expected fraction of red balls after the next (i.e. second) draw is

$$\frac{2}{3} \cdot \frac{(2+1)}{4} + \frac{1}{3} \cdot \frac{2}{4} = \frac{6+2}{12} = \frac{2}{3}$$

which is of course equal to the proportion of red balls that we had at $n = 1$.

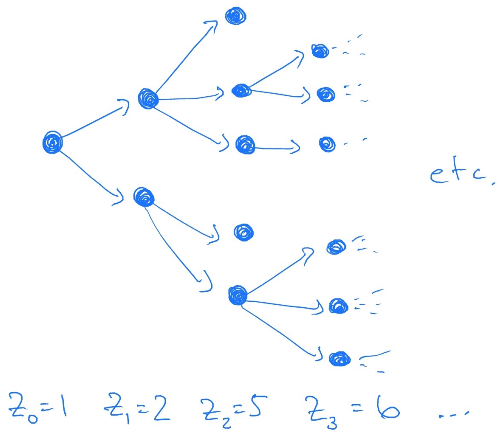
Lastly, note that it is equally likely that, on the first go, you'd pick out a black. So, starting from $n = 0$ and looking forwards, both colors have equally good chances of increasing their own numbers. In fact, if we were to pretend, right from the start, that black was red, and red was black, we would see the same urn process. This type of fairness is known as symmetry. We've seen that B_n tends to increase (because we keep adding more balls), and we can think of M_n as a way of discovering the fairness 'hiding' inside of B_n.

To sum up: in life there are different ways to think of 'fairness' – and what we need to do here is get a sense for precisely what kind of fairness martingales characterize. The fact that (M_n) is a martingale does not prevent us from (sometimes) ending up with many more red balls than black, or vice versa. It just means that, when viewed in terms of (M_n), there is no bias towards red of black inherent in the rules of the game.

4.3 A branching process

Branching processes are stochastic processes that model objects which divide up into a random number of copies of themselves. They are particularly important in mathematical biology (think of cell division, the tree of life, etc). We won't study any mathematical biology in this course, but we will look at one example of a branching process: the Bienaymé-Galton-Watson process.

The Bienaymé-Galton-Watson (BGW for short) process is parametrized by a random variable G, which is known as the *offspring distribution*. It is simplest to understand the BGW process by drawing a tree, for example:



Each dot is a 'parent', which has a random number of child dots (indicated by arrows). Each parent choses how many children it will have independently of all else, by taking a random sample of G. The BGW process is the process Z_n, where Z_n is the number of dots in generation n.

Formally, we define the BGW process as follows. Let X_i^n, where $n, i \geq 1$, be i.i.d. nonnegative integer-valued random variables with common distribution G. Define a sequence (Z_n) by $Z_0 = 1$ and

$$Z_{n+1} = \begin{cases} X_1^{n+1} + \ldots + X_{Z_n}^{n+1}, & \text{if } Z_n > 0 \\ 0, & \text{if } Z_n = 0 \end{cases} \tag{4.5}$$

Then Z is the BGW process. The random variable X_i^n represents the number of children of the i^{th} parent in the n^{th} generation.

Note that if $Z_n = 0$ for some n, then for all $m > n$ we also have $Z_m = 0$.

Remark 4.3.1 The Bienaymé-Galton-Watson is commonly just called Galton-Watson process in the literature. It takes its name from Francis Galton (a statistician and social scientist) and Henry Watson (a mathematical physicist), who in 1874 were concerned that Victorian aristocratic surnames were becoming extinct. They tried to model how many children people had, which is also how many times a surname was passed on, per family. This allowed them to use the process Z_n to predict whether a surname would die out (i.e. if $Z_n = 0$ for some n) or become widespread (i.e. $Z_n \to \infty$). Unfortunately they made mistakes in their work, and in fact it turns out the French mathematician Irénée-Jules Bienaymé had entirely solved the question in 1845.

(Since then, the BGW process has found more important uses.)

Let us assume that $G \in L^1$ and write $\mu = \mathbb{E}[G]$. Let $\mathcal{F}_n = \sigma(X_m^i\,;\, i \in \mathbb{N}, m \leq n)$. In general,

Z_n is not a martingale because

$$\begin{aligned}
\mathbb{E}[Z_{n+1}] &= \mathbb{E}\left[X_1^{n+1} + \ldots + X_{Z_n}^{n+1}\right] \\
&= \mathbb{E}\left[\sum_{k=1}^{\infty}\left(X_1^{n+1} + \ldots + X_k^{n+1}\right)\mathbb{1}\{Z_n = k\}\right] \\
&= \sum_{k=1}^{\infty}\mathbb{E}\left[\left(X_1^{n+1} + \ldots + X_k^{n+1}\right)\mathbb{1}\{Z_n = k\}\right] \\
&= \sum_{k=1}^{\infty}\mathbb{E}\left[\left(X_1^{n+1} + \ldots + X_k^{n+1}\right)\right]\mathbb{E}\left[\mathbb{1}\{Z_n = k\}\right] \\
&= \sum_{k=1}^{\infty}\left(\mathbb{E}\left[X_1^{n+1}\right] + \ldots + \mathbb{E}\left[X_k^{n+1}\right]\right)\mathbb{P}\left[Z_n = k\right] \\
&= \sum_{k=1}^{\infty} k\mu\mathbb{P}[Z_n = k] \\
&= \mu\sum_{k=1}^{\infty} k\mathbb{P}[Z_n = k] \\
&= \mu\,\mathbb{E}[Z_n]. \qquad (4.6)
\end{aligned}$$

Here, we use that the X_i^{n+1} are independent of $\mathcal{F}_n$, but Z_n (and hence also $\mathbb{1}\{Z_n = k\}$) is $\mathcal{F}_n$ measurable. We justify exchanging the infinite $\sum$ and $\mathbb{E}$ using the result of exercise (**6.8**).

From (4.6), Lemma 3.3.6 tells us that if (M_n) is a martingale that $\mathbb{E}[M_n] = \mathbb{E}[M_{n+1}]$. But, if $\mu < 1$ we see that $\mathbb{E}[Z_{n+1}] < \mathbb{E}[Z_n]$ (downwards drift) and if $\mu > 1$ then $\mathbb{E}[Z_{n+1}] > \mathbb{E}[Z_n]$ (upwards drift).

However, much like with the asymmetric random walk, we can compensate for the drift and obtain a martingale. More precisely, we will show that

$$M_n = \frac{Z_n}{\mu^n}$$

is a martingale.

We have $M_0 = 1 \in m\mathcal{F}_0$, and if $M_n \in \mathcal{F}_n$ then from (4.5) we have that $M_{n+1} \in m\mathcal{F}_{n+1}$. Hence, by induction $M_n \in \mathcal{F}_n$ for all $n \in \mathbb{N}$. From (4.6), we have $\mathbb{E}[Z_{n+1}] = \mu\mathbb{E}[Z_n]$ so as $\mathbb{E}[Z_n] = \mu^n$ for all n. Hence $\mathbb{E}[M_n] = 1$ and $M_n \in L^1$.

Lastly, we repeat the calculation that led to (4.6), but now with conditional expectation in place of $\mathbb{E}$. The first few steps are essentially the same, and we obtain

$$\begin{aligned}
\mathbb{E}[Z_{n+1} \,|\, \mathcal{F}_n] &= \sum_{k=1}^{\infty}\mathbb{E}\left[\left(X_1^{n+1} + \ldots + X_k^{n+1}\right)\mathbb{1}\{Z_n = k\} \,|\, \mathcal{F}_n\right] \\
&= \sum_{k=1}^{\infty}\mathbb{1}\{Z_n = k\}\mathbb{E}\left[X_1^{n+1} + \ldots + X_k^{n+1} \,|\, \mathcal{F}_n\right] \\
&= \sum_{k=1}^{\infty}\mathbb{1}\{Z_n = k\}\mathbb{E}\left[X_1^{n+1} + \ldots + X_k^{n+1}\right] \\
&= \sum_{k=1}^{\infty} k\mu\mathbb{1}\{Z_n = k\} \\
&= \mu\sum_{k=1}^{\infty} k\mathbb{1}\{Z_n = k\}
\end{aligned}$$

$$= \mu Z_n.$$

Here we use that Z_n is $\mathcal{F}_n$ measurable to take out what is known, and then use that X_i^{n+1} is independent of $\mathcal{F}_n$. Hence, $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n$, as required.

4.4 Other stochastic processes

The world of stochastic processes, like the physical world that they try to model, is many and varied. We can make more general kinds of random walk (and urn/branching processes) by allowing more complex rules for what should happen on each new time step. Those of you who have taken MAS2003 will have seen Poisson processes and Markov chains, which are two more important types of stochastic process. There are stochastic processes to model objects that coalesce together, objects that move around in space, objects that avoid one another, objects that repeat themselves, objects that modify themselves, etc, etc.

Most (but not quite all) types of stochastic process have connections to martingales. The reason for making these connections is that by using martingales it is possible to extract information about the behaviour of a stochastic process – we will see some examples of how this can be done in Chapters 7 and 8.

Remark 4.4.1 All the processes we have studied in this section can be represented as Markov chains with state space $\mathbb{N}$. It is possible to use the general theory of Markov chains to study these stochastic processes, but it wouldn't provide as much detail as we will obtain (in Chapters 7 and 8) using martingales.

4.5 Exercises on Chapter 4

On stochastic processes

4.1 Let $S_n = \sum_{i=1}^n X_i$ be the symmetric random walk from Section 4.1 and let $Z_n = e^{S_n}$. Show that (Z_n) is a submartingale and that

$$M_n = \left(\frac{2}{e + \frac{1}{e}}\right)^n Z_n$$

is a martingale.

4.2 Let $S_n = \sum_{i=1}^n X_i$ be the asymmetric random walk from Section 4.1, where $\mathbb{P}[X_i = 1] = p$, $\mathbb{P}[X_i = -1] = q$ and with $p > q$ and $p + q = 1$. Show that (S_n) is a submartingale and that

$$M_n = \left(\frac{q}{p}\right)^{S_n}$$

is a martingale.

4.3 Let (X_i) be a sequence of identically distributed random variables with common distribution

$$X_i = \begin{cases} a & \text{with probability } p_a \\ -b & \text{with probability } p_b = 1 - p_a. \end{cases}$$

where $0 \leq a, b$. Let $S_n = \sum_{i=1}^n X_i$. Under what conditions on a, b, p_a, p_b is (S_n) a martingale?

4.4 Let $S_n = \sum_{i=1}^n X_i$ be the symmetric random walk from Section 4.1. Show that S_n^2 is a submartingale and that $M_n = S_n^2 - n$ is a martingale.

4.5 Let (X_i) be an i.i.d. sequence of random variables such that $\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = -1] = \frac{1}{2}$. Define a stochastic process S_n by setting $S_0 = 1$ and

$$S_{n+1} = \begin{cases} S_n + X_{n+1} & \text{if } S_n > 0, \\ 1 & \text{if } S_n = 0. \end{cases}$$

That is, S_n behaves like a symmetric random walk but, whenever it becomes zero, on the next time step it is 'reflected' back to 1. Let

$$L_n = \sum_{i=0}^{n-1} \mathbb{1}\{S_i = 0\}$$

be the number of time steps, before time n, at which S_n is zero. Show that

$$\mathbb{E}[S_{n+1} \,|\, \mathcal{F}_n] = S_n + \mathbb{1}\{S_n = 0\}$$

and hence show that $S_n - L_n$ is a martingale.

4.6 Consider an urn that may contain balls of three colours: red, blue and green. Initially the urn contains one ball of each colour. Then, at each step of time $n = 1, 2, \ldots$ we draw a ball from the urn. We place the drawn ball back into the urn and add an additional ball of the same colour.

Let (M_n) be the proportion of balls that are red. Show that (M_n) is a martingale.

4.7 Let $S_n = \sum_{i=1}^n X_i$ be the symmetric random walk from Section 4.1. State, with proof, which of the following processes are martingales:

$$\text{(i) } S_n^2 + n \qquad \text{(ii) } S_n^2 + S_n - n \qquad \text{(iii) } \frac{S_n}{n}$$

Which of the above are submartingales?

Challenge questions

4.8 Let (S_n) be the symmetric random walk from Section 4.1. Prove that there is no deterministic function $f : \mathbb{N} \to \mathbb{R}$ such that $S_n^3 - f(n)$ is a martingale.

Chapter 5

The binomial model

We now return to financial mathematics. We will extend the one-period model from Chapter 1 and discover a surprising connection between arbitrage and martingales.

5.1 Arbitrage in the one-period model

Let us recall the one-period market from Section 1.2. We have two commodities, cash and stock. Cash earns interest at rate r, so:

- If we hold x units of cash at time 0, they become worth $x(1+r)$ at time 1.

At time $t = 0$, a single unit of stock is worth s units of cash. At time 1, the value of a unit of stock changes to

$$S_1 = \begin{cases} sd & \text{with probability } p_d, \\ su & \text{with probability } p_u, \end{cases}$$

where $p_u + p_d = 1$.

Note that roles of u and d are interchangeable – we would get the same model if we swapped the values of u and d (and p_u and p_d to match). So, we lose nothing by assuming that $d < u$. We also assume that all of r, p_d, p_d, d and u are strictly positive.

The price of our stock changes as follows:

- If we hold y units of stock, worth ys, at time 0, they become worth yS_1 at time 1.

Recall that we can borrow cash from the bank (provided we pay it back with interest at rate r, at some later time) and that we can borrow stock from the stockbroker (provided we give the same number of units of stock back, at some later time). Thus, x and y are allowed to be negative, with the meaning that we have borrowed.

Recall also that we use the term portfolio for the amount of cash/stock that we hold at some time. We can formalize this: A **portfolio** is a pair $h = (x, y) \in \mathbb{R}^2$, where x is the amount of cash and y is the number of (units of) of stock.

Definition 5.1.1 The **value process** or **price process** of the portfolio $h = (x, y)$ is the process V^h given by

$$V_0^h = x + ys$$
$$V_1^h = x(1+r) + yS_1.$$

We can also formalize the idea of arbitrage. A portfolio is an arbitrage if it makes money for free:

Definition 5.1.2 A portfolio $h = (x, y)$ is said to be an **arbitrage possibility** if:

$$\begin{aligned} V_0^h &= 0 \\ \mathbb{P}[V_1^h \geq 0] &= 1 \\ \mathbb{P}[V_1^h > 0] &> 0. \end{aligned}$$

We say that a market is **arbitrage free** if there do not exist any arbitrage possibilities.

It is possible to characterize exactly when the one-period market is arbitrage free. In fact, we have already done most of the work in exercise **1.3**.

Proposition 5.1.3 *The one-period market is arbitrage free if and only if $d < 1 + r < u$.*

Proof: $(\Rightarrow)$: Recall that we assume $d < u$. Hence, if $d < 1+r < u$ fails then either $1+r \leq d < u$ or $d < u \leq 1 + r$. In both cases, we will construct an arbitrage possibility.

In the case $1 + r \leq d < u$ we use the portfolio $h = (-s, 1)$ which has $V_0^h = 0$ and

$$V_1^h = -s(1 + r) + S_1 \geq s(-(1 + r) + d) \geq 0,$$

hence $\mathbb{P}[V_1^h \geq 0] = 1$. Further, with probability $p_u > 0$ we have $S_1 = su$, which means $V_1^h > s(-(1 + r) + d) \geq 0$. Hence $\mathbb{P}[V_1^h > 0] > 0$. Thus, h is an arbitrage possibility.

If $0 < d < u \leq 1 + r$ then we use the portfolio $h' = (s, -1)$, which has $V_0^{h'} = 0$ and

$$V_1^{h'} = s(1 + r) - S_1 \geq s(1 + r - u) \geq 0,$$

hence $\mathbb{P}[V_1^{h'} \geq 0] = 1$. Further, with probability $p_d > 0$ we have $S_1 = sd$, which means $V_1^{h'} > s(-(1 + r) + u) \geq 0$. Hence $\mathbb{P}[V_1^{h'} > 0] > 0$. Thus, h' is also an arbitrage possibility.

Remark 5.1.4 In both cases, at time 0 we borrow whichever commodity (cash or stock) will grow slowest in value, immediately sell it and use the proceeds to buy the other, which we know will grow faster in value. Then we wait; at time 1 we own the commodity has grown fastest in value, so we sell it, repay our debt and have some profit left over.

$(\Leftarrow)$: Now, assume that $d < 1 + r < u$. We need to show that no arbitrage is possible. To do so, we will show that if a portfolio has $V_0^h = 0$ and $V_1^h \geq 0$ then it also has $V_1^h = 0$.

So, let $h = (x, y)$ be a portfolio such that $V_0^h = 0$ and $V_1^h \geq 0$. We have

$$V_0^h = x + ys = 0.$$

The value of h at time 1 is

$$V_1^h = x(1 + r) + ysZ.$$

Using that $x = -ys$, we have

$$V_1^h = \begin{cases} ys(u - (1 + r)) & \text{if } Z = u, \\ ys(d - (1 + r)) & \text{if } Z = d. \end{cases} \tag{5.1}$$

Since $\mathbb{P}[V_1^h \geq 0] = 1$ this means that both (a) $ys(u - (1 + r)) \geq 0$ and (b) $ys(d - (1 + r)) \geq 0$. If $y < 0$ then we contradict (a) because $1+r < u$. If $y > 0$ then we contradict (b) because $d < 1+r$. So the only option left is that $y = 0$, in which case $V_0^h = V_1^h = 0$. ∎

Expectation regained

In Proposition 5.1.3 we showed that our one period model was free of arbitrage if and only if

$$d < 1 + r < u.$$

This condition is very natural: it means that sometimes the stock will outperform cash and sometimes cash will outperform the stock. Without this condition it is intuitively clear that our market would be a bad model. From that point of view, Proposition 5.1.3 is encouraging since it confirms the importance of (no) arbitrage.

However, it turns out that there is more to the condition $d < 1+r < u$, which we now explore. It is equivalent to asking that there exists $q_u, q_d \in (0,1)$ such that both

$$q_u + q_d = 1 \qquad \text{and} \qquad 1 + r = uq_u + dq_d. \tag{5.2}$$

In words, (5.2) says that $1+r$ is a weighted average of d and u, where d has weight q_d and u has weight q_u. Solving these two equations gives

$$q_u = \frac{(1+r) - d}{u - d}, \qquad q_d = \frac{u - (1+r)}{u - d}. \tag{5.3}$$

Now, here is the key: we can think of the weights q_u and q_d as probabilities. Let's pretend that we live in a different world, where a single unit of stock, worth $S_0 = s$ at time 0, changes value to become worth

$$S_1 = \begin{cases} sd & \text{with probability } q_d, \\ su & \text{with probability } q_u. \end{cases}$$

We have altered – the technical term is **tilted** – the probabilities from their old values p_d, p_u to new values q_d, q_u. Let's call this new world $\mathbb{Q}$, by which we mean that $\mathbb{Q}$ is our new probability measure: $\mathbb{Q}[S_1 = sd] = q_d$ and $\mathbb{Q}[S_1 = su] = q_u$. This is often called the **risk-neutral world**, and q_u, q_d are known as the **risk-neutral probabilities**[1].

Since $\mathbb{Q}$ is a probability measure, we can use it to take expectations. We use $\mathbb{E}^{\mathbb{P}}$ and $\mathbb{E}^{\mathbb{Q}}$ to make it clear if we are taking expectations using $\mathbb{P}$ or $\mathbb{Q}$.

We have

$$\begin{aligned}
\frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}[S_1] &= \frac{1}{1+r}\Big(su\mathbb{Q}[S_1 = su] + sd\mathbb{Q}[S_1 = sd]\Big) \\
&= \frac{1}{1+r}(s)(uq_u + dq_d) \\
&= s.
\end{aligned}$$

The price of the stock at time 0 is $S_0 = s$. To sum up, we have shown that the price S_1 of a unit of stock at time 1 satisfies

$$S_0 = \frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}[S_1]. \tag{5.4}$$

This is a formula that is very well known to economists. It gives the stock price today ($t = 0$) as the expectation under $\mathbb{Q}$ of the stock price tomorrow ($t = 1$), **discounted** by the rate $1 + r$ at which it would earn interest.

Equation (5.4) is our first example of a 'risk-neutral valuation' formula. Recall that we pointed out in Chapter 1 that we should not use $\mathbb{E}^{\mathbb{P}}$ and 'expected value' prices. A possible cause of

[1] We will discuss the reason for the name 'risk-neutral' later.

confusion is that (5.4) *does* correctly calculate the value (i.e. price) of a single unit of stock by taking an expectation. The point is that we (1) use $\mathbb{E}^{\mathbb{Q}}$ rather than $\mathbb{E}^{\mathbb{P}}$ and (2) then discount according to the interest rate. We will see, in the next section, that these two steps are the correct way to go about arbitrage free pricing in general.

Moreover, in Section 5.4 we will extend our model to have multiple time steps. Then the expectation in (5.4) will lead us to martingales.

5.2 Hedging in the one-period model

We saw in Section 1.3 that the 'no arbitrage' assumption could force *some* prices to take particular values. It is not immediately obvious if the absence of arbitrage forces a unique value for *every* price; we will show in this section that it does.

First, let us write down exactly what it is that we need to price.

Definition 5.2.1 A **contingent claim** is any random variable of the form $X = \Phi(S_1)$, where Φ is a deterministic function.

The function Φ is sometimes known as the contract function. One example of a contingent claim is a **forward contract**, in which the holder promises to buy a unit of stock at time 1 for a fixed price K, known as the **strike price**. In this case the contingent claim would be

$$\Phi(S_1) = S_1 - K,$$

the value of a unit of stock at time 1 minus the price paid for it. We will see many other examples in the course. Here is another.

Example 5.2.2 A **European call option** gives its holder the right (but not the obligation) to buy, at time 1, a single unit of stock for a fixed price K that is agreed at time 0. As for futures, K is known as the strike price.

Suppose we hold a European call option at time 1. Then, if $S_1 > K$, we could exercise our right to buy a unit of stock at price K, immediately sell the stock for S_1 and consequently earn $S_1 - K > 0$ in cash. Alternatively if $S_1 \leq K$ then our option is worthless.

Since S_1 is equal to either to either su or sd, the only interesting case is when $sd < K < su$. In this case, the contingent claim for our European call option is

$$\Phi(S_1) = \begin{cases} su - K & \text{if } S_1 = su \\ 0 & \text{if } S_1 = sd. \end{cases} \tag{5.5}$$

In the first case our right to buy is worth exercising; in the second case it is not. A simpler way to write this contingent claim is

$$\Phi(S_1) = \max(S_1 - K, 0). \tag{5.6}$$

In general, given any contract, we can work out its contingent claim. We therefore plan to find a general way of pricing contingent claims. In Section 1.3 we relied on finding specific trading strategies to determine prices (one, from the point of view of the buyer, that gave an upper bound and one, from the point of view of the seller, to give a lower bound). Our first step in this section is to find a general way of constructing trading strategies.

Definition 5.2.3 We say that a portfolio h is a **replicating portfolio** or **hedging portfolio** for the contingent claim $\Phi(S_1)$ if $V_1^h = \Phi(S_1)$.

The process of finding a replicating portfolio is known simply as replicating or hedging. The above definition means that, if we hold the portfolio h at time 0, then at time 1 it will have precisely the same value as the contingent claim $\Phi(S_1)$. Therefore, since we assume our model is free of arbitrage:

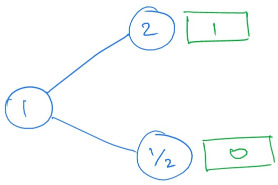
> If a contingent claim $\Phi(S_1)$ has a replicating portfolio h, then the price of the $\Phi(S_1)$ at time 0 must be equal to the value of h at time 0.

We say that a market is **complete** if every contingent claim can be replicated. Therefore, if the market is complete, we can price any contingent claim.

Example 5.2.4 Suppose that $s = 1, d = \frac{1}{2}, u = 2$ and $r = \frac{1}{4}$, and that we are looking at the contingent claim

$$\Phi(S_1) = \begin{cases} 1 & \text{if } S_1 = su, \\ 0 & \text{if } S_1 = sd. \end{cases}$$

We can represent this situation as a tree, with a branch for each possible movement of the stock, and the resulting value of our contingent claim written in a square box.



Suppose that we wish to replicate $\Phi(S_1)$. That is, we need a portfolio $h = (x, y)$ such that $V_1^h = \Phi(S_1)$:

$$(1 + \tfrac{1}{4})x + 2y = 1$$
$$(1 + \tfrac{1}{4})x + \tfrac{1}{2}y = 0.$$

This is a pair of linear equations that we can solve. The solution (which is left for you to check) is $x = \frac{-4}{15}$, $y = \frac{2}{3}$. Hence the price of our contingent claim $\Phi(S_1)$ at time 0 is $V_0^h = \frac{-4}{15} + 1 \cdot \frac{2}{3} = \frac{2}{5}$.

Let us now take an arbitrary contingent claim $\Phi(S_1)$ and see if we can replicate it. This would mean finding a portfolio h such that the value V_1^h of the portfolio at time 1 is $\Phi(S_1)$:

$$V_1^h = \begin{cases} \Phi(su) & \text{if } S_1 = su, \\ \Phi(sd) & \text{if } S_1 = sd. \end{cases}$$

By (5.1), if we write $h = (x, y)$ then we need

$$(1 + r)x + suy = \Phi(su)$$
$$(1 + r)x + sdy = \Phi(sd),$$

which is just a pair of linear equations to solve for (x, y). In matrix form,

$$\begin{pmatrix} 1+r & su \\ 1+r & sd \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \Phi(su) \\ \Phi(sd) \end{pmatrix}. \tag{5.7}$$

A unique solution exists when the determinant is non-zero, that is when $(1+r)su - (1+r)sd \neq 0$, or equivalently when $u \neq d$. So, in this case, we can find a replicating portfolio for any contingent claim.

It is an assumption of the model that $d \leq u$, so we have that our one-period model is complete if $d < u$. Therefore:

Proposition 5.2.5 *If the one-period model is arbitrage free then it is complete.*

And, in this case, we can solve (5.7) to get

$$x = \frac{1}{1+r}\frac{u\Phi(sd) - d\Phi(su)}{u-d},$$
$$y = \frac{1}{s}\frac{\Phi(su) - \Phi(sd)}{u-d}. \tag{5.8}$$

which tells us that the price of $\Phi(S_1)$ at time 0 should be

$$\begin{aligned} V_0^h &= x + sy \\ &= \frac{1}{1+r}\left(\frac{(1+r)-d}{u-d}\Phi(su) + \frac{u-(1+r)}{u-d}\Phi(sd)\right) \\ &= \frac{1}{1+r}\left(q_u\Phi(su) + q_d\Phi(sd)\right) \\ &= \frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}[\Phi(S_1)]. \end{aligned}$$

Hence, the value (and therefore, the price) of $\Phi(S_1)$ at time 0, is given by

$$V_0^h = \frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}[\Phi(S_1)]. \tag{5.9}$$

The formula (5.9) is known as the **risk-neutral valuation formula**. It says that to find the price of $\Phi(S_1)$ at time 0 we should take its expectation according to $\mathbb{Q}$, and then discount one time step worth of interest i.e. divide by $1+r$. It is a very powerful tool, since it allows us to price any contingent claim.

Note the similarity of (5.9) to (5.4). In fact, (5.4) is a special case of (5.9), namely the case where $\Phi(S_1) = S_1$ i.e. pricing the contingent claim corresponding to being given a single unit of stock.

To sum up:

Proposition 5.2.6 *Let $\Phi(S_1)$ be a contingent claim. Then the (unique) replicating portfolio $h = (x, y)$ for $\Phi(S_1)$ can be found by solving $V_1^h = \Phi(S_1)$, which can be written as a pair of linear equations:*

$$(1+r)x + suy = \Phi(su)$$
$$(1+r)x + sdy = \Phi(sd).$$

The general solution is (5.8)*. The value (and hence, the price) of $\Phi(S_1)$ at time 0 is*

$$V_0^h = \frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}\left[\Phi(S_1)\right].$$

For example, we can now both price and hedge the European call option.

Example 5.2.7 In (5.5) we found the contingent claim of a European call option with strike price $K \in (sd, su)$ to be $\Phi(S_T) = \max(S_T - K, 0)$. By the first part of Proposition 5.2.6, to find a replicating portfolio $h = (x, y)$ we must solve $V_1^h = \Phi(S_1)$, which is

$$(1+r)x + suy = su - K$$
$$(1+r)x + sdy = 0.$$

This has the solution (again, left for you to check) $x = \frac{sd(K-su)}{(1+r)(su-sd)}, y = \frac{su-K}{su-sd}$. By the second part of Proposition 5.2.6 the value of the European call option at time 0 is

$$\begin{aligned}\frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}\left[\Phi(S_1)\right] &= \frac{1}{1+r}\left(q_u(su-K)+q_d(0)\right)\\ &= \frac{1}{1+r}\frac{(1+r)-d}{u-d}(su-K).\end{aligned}$$

5.3 Types of financial derivative

A contract that specifies that buying/selling will occur, now or in the future, is known as a **financial derivative**, or simply derivative. Financial derivatives that give a choice between two options are often known simply as **options**.

Here we collect together the three types of financial derivatives that we have mentioned in previous sections. As before, we use the term **strike price** to refer to a fixed price K that is agreed at time 0 (and paid at time 1).

- A **forward** or **forward contract** is the obligation to buy (or sell) a single unit of stock at time 1 for a strike price K.
- A **European call option** is the right, but not the obligation, to buy a single unit of stock at time 1 for a strike price K.
- A **European put option** is the right, but not the obligation, to sell a single unit of stock at time 1 for a strike price K.

You are expected to remember these definitions! We will often use them in our examples.

There are many other types of financial derivative; we'll mention some more examples later in the course, in Section 17.2.

5.4 The binomial model

Let us step back and examine our progress, for a moment. We now know about as much about one-period model as there is to know. It is time to move onto to a more complicated (and more realistic) model. The one-period model is unsatisfactory in two main respects:

1. The one-period model has only a single step of time.
2. The stock price process (S_t) is too simplistic.

We'll start to address the first of these points now. The second point waits until the second semester of the course.

Adding multiple time steps to our model will make use of the theory we developed in Chapters 2 and 3. It will also reveal a surprising connection between arbitrage and martingales.

The **binomial model** has time points $t = 0, 1, 2, \ldots, T$. Inside each time step, we have a single step of the one-period model. This means that cash earns interest at rate r:

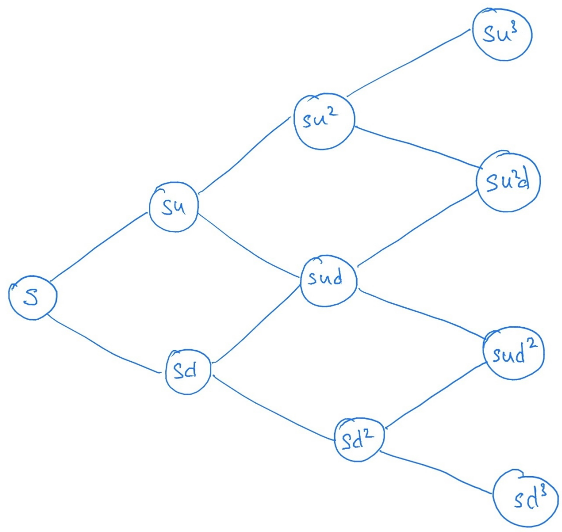
- If we hold x units of cash at time t, it will become worth $x(1+r)$ (in cash) at time $t+1$.

For our stock, we'll have to think a little harder. In a single time step, the value of our stock is multiplied by a random variable Z with distribution $\mathbb{P}[Z = u] = p_u$, $\mathbb{P}[Z = d] = p_d$. We now have several time steps. For each time step we'll use a new *independent* Z. So, let $(Z_t)_{t=1}^T$ be a sequence of i.i.d. random variables each with the distribution of Z.

- The value of a single unit of stock at time t is given by

$$S_0 = s,$$
$$S_t = Z_t S_{t-1}.$$

We can illustrate the process (S_t) using a tree-like diagram:



Note that the tree is recombining, in the sense that a move up (by u) followed by a move down (by d) has the same outcome as a move down followed by a move up. It's like a random walk, except we multiply instead of add (recall exercise **4.1**).

Remark 5.4.1 The one-period model is simply the $T = 1$ case of the binomial model. Both models are summarized on the formula sheet, see Appendix B.

5.5 Portfolios, arbitrage and martingales

Since we now have multiple time steps, we can exchange cash for stock (and vice versa) at all times $t = 0, 1, \dots, T-1$. We need to expand our idea of a portfolio to allow for this.

The filtration corresponding to the information available to a buyer/seller in the binomial model is

$$\mathcal{F}_t = \sigma(Z_1, Z_2, \dots, Z_t).$$

In words, the information in $\mathcal{F}_t$ contains changes in the stock price up to and including at time t. This means that, $S_0, S_1, \dots, S_t$ are all $\mathcal{F}_t$ measurable, but S_{t+1} is not $\mathcal{F}_t$ measurable.

When we choose how much stock/cash to buy/sell at time $t-1$, we do so without knowing how the stock price will change during $t-1 \mapsto t$. So we must do so using information only from $\mathcal{F}_{t-1}$.

We now have enough terminology to define the strategies that are available to participants in the binomial market.

Definition 5.5.1 A **portfolio strategy** is a stochastic process

$$h_t = (x_t, y_t)$$

for $t = 1, 2, \dots, T$, such that h_t is $\mathcal{F}_{t-1}$ measurable.

The interpretation is that x_t is the amount of cash, and y_t the amount of stock, that we hold during the time step $t-1 \mapsto t$. We make our choice of how much cash and stock to hold during $t-1 \mapsto t$ based on knowing the value of $S_0, S_1, \dots, S_{t-1}$, but *without* knowing S_t. This is realistic.

Definition 5.5.2 The **value process** of the portfolio strategy $h = (h_t)_{t=1}^T$ is the stochastic process $(V_t^h)_{t=0}^T$ given by

$$\begin{aligned} V_0^h &= x_1 + y_1 S_0, \\ V_t^h &= x_t(1+r) + y_t S_t, \end{aligned}$$

for $t = 1, 2, \dots, T$.

At $t = 0$, V_0^h is the value of the portfolio h_1. For $t \geq 1$, V_t^h is the value of the portfolio h_t at time t, after the change in value of cash/stock that occurs during $t-1 \mapsto t$. The value process V_t^h is $\mathcal{F}_t$ measurable but it is not $\mathcal{F}_{t-1}$ measurable.

We will be especially interested in portfolio strategies that require an initial investment at time 0 but, at later times $t \geq 1, 2, \dots, T-1$, any changes in the amount of stock/cash held will pay for itself. We capture such portfolio strategies in the following definition.

Definition 5.5.3 A portfolio strategy $h_t = (x_t, y_t)$ is said to be **self-financing** if

$$V_t^h = x_{t+1} + y_{t+1} S_t.$$

for $t = 0, 1, 2, \dots, T$.

This means that the value of the portfolio at time t is equal to the value (at time t) of the stock/cash that is held in between times $t \mapsto t+1$. In other words, in a self-financing portfolio at the times $t = 1, 2, \dots$ we can swap our stocks for cash (and vice versa) according to whatever the stock price turns out to be, but that is all we can do.

Lastly, our idea of arbitrage must also be upgraded to handle multiple time steps.

Definition 5.5.4 We say that a portfolio strategy (h_t) is an **arbitrage possibility** if it is self-financing and satisfies

$$\begin{aligned} V_0^h &= 0 \\ \mathbb{P}[V_T^h \geq 0] &= 1. \\ \mathbb{P}[V_T^h > 0] &> 0. \end{aligned}$$

In words, an arbitrage possibility requires that we invest nothing at times $t = 0, 1, \ldots, T-1$, but which gives us a positive probability of earning something at time T, with no risk at all of actually losing money.

It's natural to ask when the binomial model is arbitrage free. Happily, the condition turns out to be the same as for the one-period model.

Proposition 5.5.5 *The binomial model is arbitrage free if and only if $d < 1 + r < u$.*

The proof is quite similar to the argument for the one-period model, but involves more technical calculations and (for this reason) we don't include it as part of the course.

Recall the risk-neutral probabilities from (5.3). In the one-period model, we use them to define the **risk-neutral** world $\mathbb{Q}$, in which on each time step the stock price moves up (by u) with probability q_u, or down (by d) with probability q_d. This provides a connection to martingales:

Proposition 5.5.6 *If $d < 1 + r < u$, then under the probability measure $\mathbb{Q}$, the process*

$$M_t = \frac{1}{(1+r)^t} S_t$$

is a martingale, with respect to the filtration $(\mathcal{F}_t)$.

PROOF: We have commented above that $S_t \in m\mathcal{F}_t$, and we also have $d^t S_0 \leq S_t \leq u^t S_0$, so S_t is bounded and hence $S_t \in L^1$. Hence also $M_t \in m\mathcal{F}_t$ and $M_t \in L^1$. It remains to show that

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}[M_{t+1} \,|\, \mathcal{F}_t] &= \mathbb{E}^{\mathbb{Q}}\left[M_{t+1}\mathbb{1}_{\{Z_{t+1}=u\}} + M_{t+1}\mathbb{1}_{\{Z_{t+1}=d\}} \,|\, \mathcal{F}_t\right] \\ &= \mathbb{E}^{\mathbb{Q}}\left[\frac{uS_t}{(1+r)^{t+1}}\mathbb{1}_{\{Z_{t+1}=u\}} + \frac{dS_t}{(1+r)^{t+1}}\mathbb{1}_{\{Z_{t+1}=d\}} \,|\, \mathcal{F}_t\right] \\ &= \frac{S_t}{(1+r)^{t+1}}\left(u\mathbb{E}^{\mathbb{Q}}\left[\mathbb{1}_{\{Z_{t+1}=u\}} \,|\, \mathcal{F}_t\right] + d\mathbb{E}^{\mathbb{Q}}\left[\mathbb{1}_{\{Z_{t+1}=d\}} \,|\, \mathcal{F}_t\right]\right) \\ &= \frac{S_t}{(1+r)^{t+1}}\left(u\mathbb{E}^{\mathbb{Q}}\left[\mathbb{1}_{\{Z_{t+1}=u\}}\right] + d\mathbb{E}^{\mathbb{Q}}\left[\mathbb{1}_{\{Z_{t+1}=d\}}\right]\right) \\ &= \frac{S_t}{(1+r)^{t+1}}\left(u\mathbb{Q}[Z_{t+1}=u] + d\mathbb{Q}\left[Z_{t+1}=d\right]\right) \\ &= \frac{S_t}{(1+r)^{t+1}}(uq_u + dq_d) \\ &= \frac{S_t}{(1+r)^{t+1}}(1+r) \\ &= M_t. \end{aligned}$$

Here, from the second to third line we take out what is known, using that $S_t \in m\mathcal{F}_t$. To deduce the third line we use linearity, and to deduce the fourth line we use that Z_{t+1} is independent of $\mathcal{F}_t$. Lastly, we recall from (5.2) that $uq_u + dq_d = 1 + r$. Hence, (M_t) is a martingale with respect to the filtration $\mathcal{F}_t$, in the risk-neutral world $\mathbb{Q}$. ■

Remark 5.5.7 Using Lemma 3.3.6 we have $\mathbb{E}^{\mathbb{Q}}[M_0] = \mathbb{E}^{\mathbb{Q}}[M_1]$, which states that $S_0 = \frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}[S_1]$. This is precisely (5.4).

5.6 Hedging

We can adapt the derivatives from Section 5.3 to the binomial model, by simply replacing time 1 with time T. For example, in the binomial model a forward contract is the obligation to buy a single unit of stock at time T for a strike price K that is agreed at time 0.

Definition 5.6.1 A **contingent claim** is a random variable of the form $\Phi(S_T)$, where $\Phi : \mathbb{R} \to \mathbb{R}$ is a deterministic function.

For a forward contract, the contingent claim would be $\Phi(S_T) = S_T - K$.

Definition 5.6.2 We say that a portfolio strategy $h = (h_t)_{t=1}^T$ is a **replicating portfolio** or **hedging strategy** for the contingent claim $\Phi(S_T)$ if h is self-financing and $V_T^h = \Phi(S_T)$.

These match the definitions for the one-period model, except we now care about the value of the asset at time T (instead of time 1). We will shortly look at how to find replicating portfolios.

As in the one-period model, the binomial model is said to be **complete** if every contingent claim can be replicated. Further, as in the one-period model, if the binomial model is free of arbitrage then it is also complete. With this in mind, for the rest of this section we assume that

$$d < 1 + r < u.$$

Lastly, as in the one-period model, our assumption that there is no arbitrage means that:

> If a contingent claim $\Phi(S_T)$ has a replicating portfolio $h = (h_t)_{t=1}^T$, then the price of the $\Phi(S_T)$ at time 0 must be equal to the value of h_0.

Now, let us end this chapter by showing how to compute prices and replicating portfolios in the binomial model. We already know how to do this in the one-period model, see Example 5.2.4. We could do it in full generality (as we did in (5.7) for the one-period model) but this would involve lots of indices and look rather messy. Instead, we'll work through a practical example that makes the general strategy clear.

Let us take $T = 3$ and set

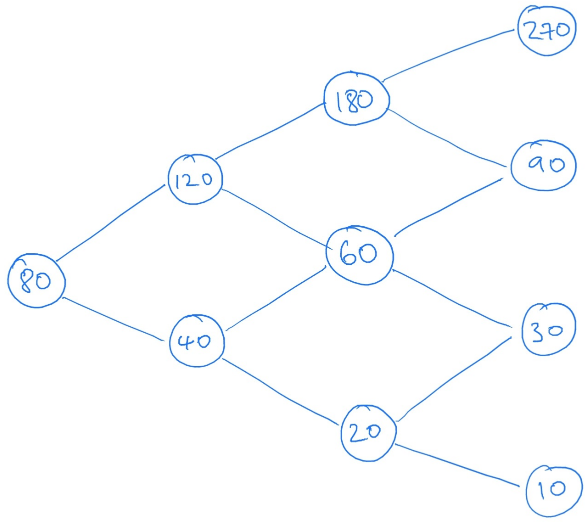
$$S_0 = 80, \quad u = 1.5, \quad d = 0.5, \quad p_u = 0.6, \quad p_d = 0.4.$$

To make the calculations easier, we'll also take our interest rate to be $r = 0$. We'll price a European call option with strike price $K = 80$. The contingent claim for this option, which is

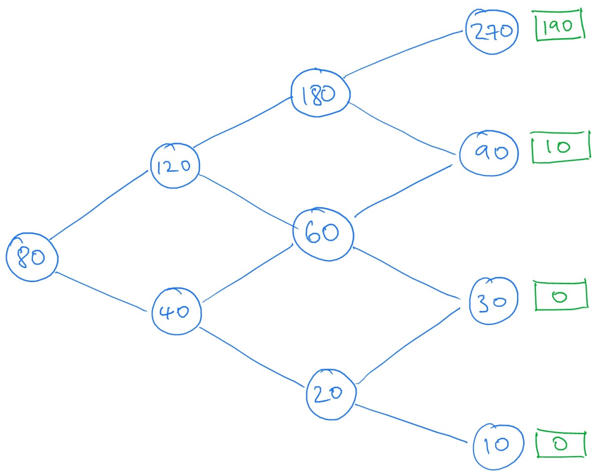
$$\Phi(S_T) = \max(S_T - K, 0). \tag{5.10}$$

STEP 1 is to work our the risk-neutral probabilities. From (5.3), these are $q_u = \frac{1+0-0.5}{1.5-0.5} = 0.5$ and $q_d = 1 - q_u = 0.5$.

STEP 2 is to write down the tree of possible values that the stock can take during time $t = 0, 1, 2, 3$. This looks like



We then work out, at each of the nodes corresponding to time $T = 3$, what the value of our contingent claim (5.10) would be if this node were reached. We write these values in square boxes:



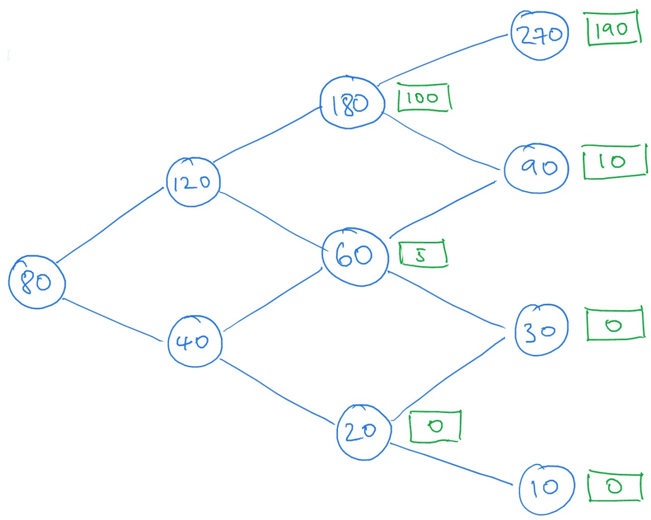
We now come to **STEP 3**, the key idea. Suppose we are sitting in one of the nodes at time $t = 2$, which we think of as the 'current' node. For example suppose we are at the uppermost node (labelled 180, the 'current' value of the stock). Looking forwards one step of time we can see that, if the stock price goes up our option is worth 190, whereas if the stock price goes down our option is worth 10. What we are seeing here is (an instance of) the one-period model! With contingent claim

$$\widetilde{\Phi}(su) = 190, \quad \widetilde{\Phi}(sd) = 10.$$

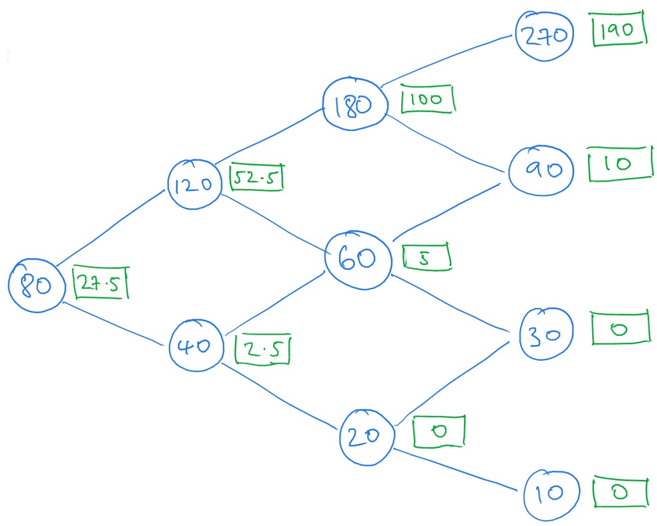
So, using the one-period risk-neutral valuation formula from Proposition 5.2.6 the value of our call option at our current node is

$$\frac{1}{1+0}(190 \cdot 0.5 + 10 \cdot 0.5) = 100.$$

We could apply the same logic to any of the nodes corresponding to time $t = 2$, and compute the value of our call option at that node:



If we now imagine ourselves sitting in one of the nodes at time $t = 1$, and look forwards one step in time, we again find ourselves faced with an instance of the one-period model. This allows us to compute the value of our call option at the $t = 1$ nodes; take for example the node labelled by 40 which, one step into the future, sees the contingent claim $\widetilde{\Phi}(su) = 5, \widetilde{\Phi}(sd) = 0$ and using (5.9) gives the value of the call option at this node as $\frac{1}{1+0}(5 \times 0.5 + 0 \times 0.5) = 2.5$. Repeating the procedure on the other $t = 1$ node, and then also on the single $t = 0$ node gives us



Therefore, the value (i.e. the price) of our call option at time $t = 0$ is 27.5.

Although we have computed the price, we haven't yet computed a replicating portfolio, which is **STEP 4**. We could do it by solving lots of linear equations for our one-period models, as in Example 5.2.4, but since we have several steps a quicker way is to apply Proposition 5.2.6 and use the general formula we found in (5.8).

Starting at time $t = 0$, to replicate the contingent claim $\widetilde{\Phi}(su) = 52.5$ and $\widetilde{\Phi}(sd) = 2.5$ at time $t = 1$, equation (5.8) tells us that we want the portfolio

$$x_1 = \frac{1}{1+0}\frac{1.5 \cdot 2.5 - 0.5 \cdot 52.5}{1.5 - 0.5} = -22.5, \quad y_1 = \frac{1}{80}\frac{52.5 - 2.5}{1.5 - 0.5} = \frac{5}{8}.$$

The value of this portfolio at time 0 is

$$x_1 + 80y_1 = -22.5 + 80 \cdot \frac{5}{8} = 27.5$$

which is equal to the initial value of our call option.

We can then carry on forwards. For example, if the stock went up in between $t = 0$ and $t = 1$, then at time $t = 1$ we would be sitting in the node for $S_1 = 120$, labelled simply 120. Our portfolio (x_1, y_1) is now worth

$$x_1(1 + 0) + y_1 \cdot 120 = -22.5 + 120 \cdot \frac{5}{8} = 52.5,$$

equal to what is now the value of our call option. We use (5.8) again to calculate the portfolio we want to hold during time $1 \mapsto 2$, this time with $\widetilde{\Phi}(su) = 100$ and $\widetilde{\Phi}(sd) = 5$, giving $x_2 = -42.5$ and $y_2 = \frac{95}{120}$. You can check that the current value of the portfolio (x_2, y_2) is 52.5.

Next, suppose the stock price falls between $t = 1$ and $t = 2$, so our next node is $S_2 = 60$. Our portfolio (x_2, y_2) now becomes worth

$$x_2(1 + 0) + y_2 \cdot 60 = -42.5 + \frac{95}{120} \cdot 60 = 5,$$

again equal to the value our of call option. For the final step, we must replicate the contingent claim $\widetilde{\Phi}(su) = 10$, $\widetilde{\Phi}(sd) = 0$, which (5.8) tells us is done using $x_3 = -5$ and $y_3 = \frac{1}{6}$. Again, you can check that the current value of this portfolio is 5.

Lastly, the stock price rises again to $S_3 = 90$. Our portfolio becomes worth

$$x_3(1 + 0) + y_3 \cdot 90 = -5 + \frac{1}{6} \cdot 90 = 10,$$

equal to the payoff from our call option.

To sum up, using (5.8) we can work out which portfolio we would want to hold, at each possible outcome of the stock changing value. At all times we would be holding a portfolio with current value equal to the current value of the call option. Therefore, this gives a self-financing portfolio strategy that replicates $\Phi(S_T)$.

5.7 Exercises on Chapter 5

All questions use the notation u, d, p_u, p_d, s and r, which has been used throughout this chapter. In all questions we assume that the models are arbitrage free and complete: $d < 1 + r < u$.

On the one-period model

5.1 Suppose that we hold the portfolio $(1, 3)$ at time 0. What is the value of this portfolio at time 1?

5.2 Find portfolios that replicate the following contingent claims.

(a) $\Phi(S_1) = 1$

(b) $\Phi(S_1) = \begin{cases} 3 & \text{if } S_1 = su, \\ 1 & \text{if } S_1 = sd. \end{cases}$.

Hence, write down the values of these contingent claims at time 0.

5.3 Find the contingent claims $\Phi(S_1)$ for the following derivatives.

(a) A contract in which the holder promises to buy two units of stock at time $t = 1$, each for strike price K.

(b) A European put option with strike price $K \in (sd, su)$ (see Section 5.3).

(c) A contract in which we promise that, if $S_1 = su$, we will sell one unit of stock at time $t = 1$ for strike price $K \in (sd, su)$ (and otherwise, if $S_1 = sd$ we do nothing).

(d) Holding both the contracts in (b) and (c) at once.

5.4 Let Π_t^{call} and Π_t^{put} be the price of European call and put options, both with the same strike price $K \in (sd, su)$, at times $t = 0, 1$.

(a) Write down formulae for Π_0^{call} and Π_0^{put}.

(b) Show that $\Pi_0^{call} - \Pi_0^{put} = s - \frac{K}{1+r}$.

On the binomial model

5.5 Write down the contingent claim of a European call option (that matures at time T).

5.6 Let $T = 2$ and let the initial value of a single unit of stock be $S_0 = 100$. Suppose that $p_u = 0.25$ and $p_d = 0.75$, that $u = 2.0$ and $d = 0.5$, and that $r = 0.25$. Draw out, in a tree-like diagram, the possible values of the stock price at times $t = 0, 1, 2$. Find the price, at time 0, of a European put option with strike price $K = 100$.

Suppose instead that $p_u = 0.1$ and $p_d = 0.9$. Does this change the value of our put option?

5.7 Let $T = 2$ and let the initial value of a single unit of stock be $S_0 = 120$. Suppose that $p_u = 0.5$ and $p_d = 0.5$, that $u = 1.5$ and $d = 0.5$, and that $r = 0.0$. Draw out, in a tree-like diagram, the possible values of the stock price at times $t = 0, 1, 2$. Annotate your tree to show a hedging strategy for a European call option with strike price $K = 60$. Hence, write down the value of this option at time 0.

5.8 Let $T = 2$ and let the initial value of a single unit of stock be $S_0 = 480$. Suppose that $p_u = 0.5$ and $p_d = 0.5$, that $u = 1.5$ and $d = 0.75$, and that $r = 0$. Draw out, in a tree-like diagram, the possible values of the stock price at times $t = 0, 1, 2$. Annotate your tree to show a hedging strategy for a European call option with strike price $K = 60$. Hence, write down the value of this option at time 0.

Comment on the values obtained for the hedging portfolios.

5.9 Recall that $(S_t)_{t=1}^T$ is the price of a single unit of stock.

(a) Find a condition on p_u, p_d, u, d that is equivalent to saying that S_t is a martingale under $\mathbb{P}$.

(b) When is $M_t = \log S_t$ is a martingale under $\mathbb{P}$?

Challenge questions

5.10 Write a computer program (in a language of your choice) that carries out the pricing algorithm for the binomial model, for a general number n of time-steps.

Chapter 6

Convergence of random variables

A real number is a simple object; it takes a single value. As such, if a_n is a sequence of real numbers, $\lim_{n\to\infty} a_n = a$, means that the value of a_n converges to the value of a.

Random variables are more complicated objects. They take many different values, with different probabilities. Consequently, if $X_1, X_2, \ldots$ and X are random variables, there are many different ways in which we can try to make sense of the idea that $X_n \to X$. They are called *modes* of convergence, and are the focus of this chapter.

Convergence of random variables sits at the heart of all sophisticated stochastic modelling. Crucially, it provides a way to approximate one random variable with another (since if $X_n \to X$ then we may hope that $X_n \approx X$ for large n), which is particularly helpful if it is possible to approximate a complex model X_n with a relatively simple random variable X. We will explore this theme further in later chapters.

6.1 Modes of convergence

We say:

- $X_n \stackrel{d}{\to} X$, known as convergence **in distribution**, if for every $x \in \mathbb{R}$ at which $\mathbb{P}[X = x] = 0$,
$$\lim_{n\to\infty} \mathbb{P}[X_n \le x] = \mathbb{P}[X \le x].$$
- $X_n \stackrel{\mathbb{P}}{\to} X$, known as convergence **in probability**, if given any $a > 0$,
$$\lim_{n\to\infty} \mathbb{P}[|X_n - X| > a] = 0.$$
- $X_n \stackrel{a.s.}{\to} X$, known as **almost sure** convergence, if
$$\mathbb{P}\left[X_n \to X \text{ as } n \to \infty\right] = 1.$$
- $X_n \stackrel{L^p}{\to} X$, known as **convergence in** L^p, if
$$\mathbb{E}\left[|X_n - X|^p\right] \to 0 \text{ as } n \to \infty.$$

Here, $p \ge 1$ is a real number. We will be interested in the cases $p = 1$ and $p = 2$. The case $p = 2$ is sometimes known as convergence in mean square. Note that these four definitions also appear on the formula sheet, in Appendix B.

It is common for random variables to converge in some modes but not others, as the following example shows.

Example 6.1.1 Let U be a uniform random variable on $[0, 1]$ and set
$$X_n = n^2 \mathbb{1}\{U < 1/n\} = \begin{cases} n^2 & \text{if } U < 1/n, \\ 0 & \text{otherwise.} \end{cases}$$

Our candidate limit is $X = 0$, the random variable that takes the deterministic value 0. We'll check each of the types of convergence in turn.

- For convergence in distribution, we note that $\mathbb{P}[X \le x] = \left\{\begin{smallmatrix} 0 & \text{if } x<0 \\ 1 & \text{if } x\ge 0 \end{smallmatrix}\right.$. We consider these two cases:
  1. Firstly, if $x < 0$ then $\mathbb{P}[X_n \le x] = 0$ so $\mathbb{P}[X_n \le x] \to 0$.
  2. Secondly, consider $x \ge 0$. By definition $\mathbb{P}[X_n = 0] = 1 - \frac{1}{n}$, so we have that $1 - \frac{1}{n} = \mathbb{P}[X_n = 0] \le \mathbb{P}[X \le x] \le 1$, and the sandwich rule tells us that $\mathbb{P}[X_n \le x] \to 1$.

  Hence, $\mathbb{P}[X_n \le x] \to \mathbb{P}[X \le x]$ in both cases, which means that $X_n \stackrel{d}{\to} X$.
- For any $0 < a \le n^2$ we have $\mathbb{P}[|X_n - 0| > a] = \mathbb{P}[X_n > a] \le \mathbb{P}[X_n = n^2] = \frac{1}{n}$, so as $n \to \infty$ we have $\mathbb{P}[|X_n - 0| > a] \to 0$, which means that we do have $X_n \stackrel{\mathbb{P}}{\to} 0$.
- If $X_m = 0$ for some $m \in \mathbb{N}$ then $X_n = 0$ for all $n \ge m$, which implies that $X_n \to 0$ as $n \to \infty$. So, we have
$$\mathbb{P}\left[\lim_{n\to\infty} X_n = 0\right] \ge \mathbb{P}[X_m = 0] = 1 - \frac{1}{m}.$$
  Since this is true for any $m \in \mathbb{N}$, we have $\mathbb{P}[\lim_{n\to\infty} X_n = 0] = 1$, that is $X_n \stackrel{a.s.}{\to} 0$.

- Lastly, $\mathbb{E}[|X_n - 0|] = \mathbb{E}[X_n] = n^2 \frac{1}{n} = n$, which does not tend to 0 as $n \to \infty$. So X_n does not converge to 0 in L^1.

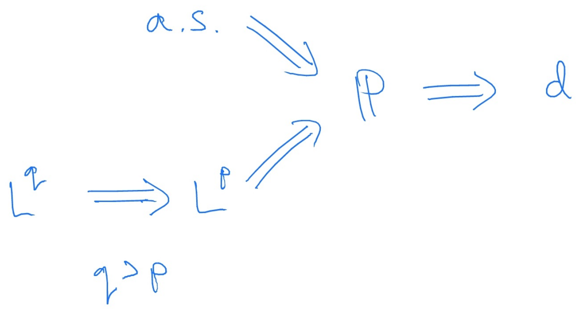
As we might hope, there are relationships between the different modes of convergence, which are useful to remember.

Lemma 6.1.2 *Let X_n, X be random variables.*

1. *If $X_n \stackrel{\mathbb{P}}{\to} X$ then $X_n \stackrel{d}{\to} X$.*
2. *If $X_n \stackrel{a.s.}{\to} X$ then $X_n \stackrel{\mathbb{P}}{\to} X$.*
3. *If $X_n \stackrel{L^p}{\to} X$ then $X_n \stackrel{\mathbb{P}}{\to} X$.*
4. *Let $1 \le p < q$. If $X_n \stackrel{L^q}{\to} X$ then $X_n \stackrel{L^p}{\to} X$.*

In all other cases (i.e. that are not automatically implied by the above), convergence in one mode does not imply convergence in another.

The proofs are not part of our course (they are part of MAS31002/61022). We can summarise Lemma 6.1.2 with a diagram:



Remark 6.1.3 For convergence of real numbers, it was shown in MAS221 that if $a_n \to a$ and $a_n \to b$ then $a = b$, which is known as uniqueness of limits. For random variables, the situation is a little more complicated: if $X_n \stackrel{\mathbb{P}}{\to} X$ and $X_n \stackrel{\mathbb{P}}{\to} Y$ then $X = Y$ almost surely. By Lemma 6.1.2, this result also applies to $\stackrel{L^p}{\to}$ and $\stackrel{a.s.}{\to}$. However, if we have only $X_n \stackrel{d}{\to} X$ and $X_n \stackrel{d}{\to} Y$ then we can only conclude that X and Y have the same distribution function. Proving these facts is one of the challenge exercises, **6.6**.

6.2 The monotone convergence theorem

A natural question to ask is, when does $\mathbb{E}[X_n] \to \mathbb{E}[X]$? This is not a mode of convergence, but simply a practical question.

We are interested (for use later on in the course) to ask when almost sure convergence implies that $\mathbb{E}[X_n] \to \mathbb{E}[X]$. As we can see from Example 6.1.1, in general it does not. We need some extra conditions:

Theorem 6.2.1 (Monotone Convergence Theorem) *Let (X_n) be a sequence of random variables and suppose that:*

1. *$X_{n+1} \geq X_n$, almost surely, for all n.*
2. *$X_n \geq 0$, almost surely, for all n.*

Then, there exists a random variable X such that $X_n \stackrel{a.s.}{\to} X$. Further, $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

The first claim of the theorem, that the sequence X_n converges almost surely, is true because by property *1* the sequence X_n is increasing, almost surely, and increasing sequences of real numbers converge[1]

Since limits preserve weak inequalities and $X_n \geq 0$ we have $X \geq 0$, almost surely. However, we should not forget that $\mathbb{E}[X]$ might be equal to $+\infty$.

You can think of Theorem 6.2.1 as a stochastic equivalent of the fact that increasing real valued sequences converge (in $\mathbb{R}$ if they are bounded, and to $+\infty$ if they are not bounded). This is usually helpful when applying the theorem, such as in the following example.

Let (X_n) be a sequence of independent random variables, with distribution given by

$$X_i = \begin{cases} 2^{-i} & \text{with probability } \frac{1}{2} \\ 0 & \text{with probability } \frac{1}{2}. \end{cases}$$

Let $Y_n = \sum_{i=1}^n X_i$. Then (Y_n) is an increasing sequence, almost surely, and hence converges almost surely to the limit $Y = \sum_{i=1}^\infty X_i$.

Since also $Y_n \geq 0$, we can apply the monotone convergence theorem to (Y_n) and deduce that $\mathbb{E}[Y_n] \to \mathbb{E}[Y]$. By linearity of $\mathbb{E}$, and geometric summation, we have that

$$\mathbb{E}[Y_n] = \sum_{i=1}^n \mathbb{E}[Y_i] = \sum_{i=1}^n \frac{1}{2}\frac{1}{2^i} = \frac{1}{2}\frac{\frac{1}{2} - (\frac{1}{2})^{n+1}}{1 - \frac{1}{2}}$$

This converges to $\frac{1}{2}$ as $n \to \infty$, so we deduce that $\mathbb{E}[Y] = \frac{1}{2}$. We'll investigate this example further in exercise 6.5. In fact, X_i corresponds to the i^{th} digit in the binary expansion of Y.

Remark 6.2.2 (Δ) Those of you taking the MAS61023 version of the course should now begin your independent reading, starting in Chapter 8. It begins with a convergence theorem similar to Theorem 6.2.1, but which applies to non-monotone sequences.

[1] ($\oslash$) More precisely: we have $\mathbb{P}[\lim_{n\to\infty} X_n(\omega) \text{ exists}] = 1$ and for ω in this set we can define $X(\omega) = \lim_{n\to\infty} X_n(\omega)$. We don't care about ω for which the limit doesn't exist, because this has probability zero! We could set $X(\omega) = 0$ in such cases, and it won't affect any probabilities involving X. See MAS31002/61022 for details.

6.3 Exercises on Chapter 6

On convergence of random variables

6.1 Let (X_n) be a sequence of independent random variables such that

$$X_n = \begin{cases} 2^{-n} & \text{with probability } \frac{1}{2} \\ 0 & \text{with probability } \frac{1}{2}. \end{cases}$$

Show that $X_n \to 0$ in L^1 and almost surely. Deduce that also $X_n \to 0$ in probability and in distribution.

6.2 Let X_n, X be random variables.

(a) Suppose that $X_n \stackrel{L^1}{\to} X$ as $n \to \infty$. Show that $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

(b) Give an example where $\mathbb{E}[X_n] \to \mathbb{E}[X]$ but X_n does not converge to X in L^1.

6.3 Let U be a random variable such that $\mathbb{P}[U=0] = \mathbb{P}[U=1] = \mathbb{P}[U=2] = \frac{1}{3}$. Let X_n and X be given by

$$X_n = \begin{cases} 1 + \frac{1}{n} & \text{if } U = 0 \\ 1 - \frac{1}{n} & \text{if } U = 1 \\ 0 & \text{if } U = 2, \end{cases} \qquad X = \begin{cases} 1 & \text{if } U \in \{0,1\} \\ 0 & \text{if } U = 2. \end{cases}$$

Show that $X_n \to X$ both almost surely and in L^1. Deduce that also $X_n \to X$ in probability and in distribution.

6.4 Let X_1 be a random variable with distribution given by $\mathbb{P}[X_1 = 1] = \mathbb{P}[X_1 = 0] = \frac{1}{2}$. Set $X_n = X_1$ for all $n \geq 2$. Set $Y = 1 - X_1$. Show that $X_n \to Y$ in distribution, but not in probability.

6.5 Let (X_n) be the sequence of random variables from **6.1**. Define $Y_n = X_1 + X_2 + \ldots + X_n$.

(a) Show that, for all $\omega \in \Omega$, the sequence $Y_n(\omega)$ is increasing and bounded.

(b) Deduce that there exists a random variable Y such that $Y_n \stackrel{a.s.}{\to} Y$.

(c) Write down the distribution of Y_1, Y_2 and Y_3.

(d) Suggest why we might guess that Y has a uniform distribution on $[0,1]$.

(e) Prove that Y_n has a uniform distribution on $\{k2^{-n}\,;\, k = 0, 1, \ldots, 2^n - 1\}$.

(f) Prove that Y has a uniform distribution on $[0,1]$.

On the monotone convergence theorem

6.6 Let Y_n be a sequence of random variables such that $Y_{n+1} \leq Y_n \leq 0$, almost surely, for all n. Show that there exists a random variable Y such that $Y_n \stackrel{a.s.}{\to} Y$ and $\mathbb{E}[Y_n] \to \mathbb{E}[Y]$.

6.7 Let X be a random variable such that $X \geq 0$ and $\mathbb{P}[X < \infty] = 1$. Define

$$X_n = \begin{cases} X & \text{if } X \leq n \\ 0 & \text{otherwise.} \end{cases}$$

Equivalently, $X_n = X\mathbb{1}\{X \leq n\}$. Show that $\mathbb{E}[X_n] \to \mathbb{E}[X]$ as $n \to \infty$.

6.8 Let (X_n) be a sequence of random variables.

(a) Explain briefly why $\mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}[X_i]$ follows from the linearity of $\mathbb{E}$, for $n \in \mathbb{N}$. Explain briefly why linearity alone does not allow us to deduce the same equation with $n = \infty$.

(b) Suppose that $X_n \geq 0$ almost surely, for all n. Show that

$$\mathbb{E}\left[\sum_{i=1}^\infty X_i\right] = \sum_{i=1}^\infty \mathbb{E}[X_i]. \tag{6.1}$$

(c) Suppose, instead, that the X_i are independent and that $\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = -1] = \frac{1}{2}$. Explain briefly why (6.1) fails to hold, in this case.

Challenge questions

6.6 Let (X_n) be a sequence of random variables, and let X and Y be random variables.

(a) Show that if $X_n \stackrel{d}{\to} X$ and $X_n \stackrel{d}{\to} Y$ then X and Y have the same distribution.

(b) Show that if $X_n \stackrel{\mathbb{P}}{\to} X$ and $X_n \stackrel{\mathbb{P}}{\to} Y$ then $X = Y$ almost surely.

6.7 Let (X_n) be a sequence of independent random variables such that $\mathbb{P}[X_n = 1] = \mathbb{P}[X_n = 0] = \frac{1}{2}$. Show that (X_n) does not converge in probability and deduce that (X_n) also does not converge in L^1, or almost surely. Does X_n converge in distribution?

Chapter 7

Stochastic processes and martingale theory

In this section we introduce two important results from the theory of martingales, namely the 'martingale transform' and the 'martingale convergence theorem'. We use these results to analyse the behaviour of stochastic processes, including those from Chapter 4 (random walks, urns, branching processes) and also the gambling game Roulette.

Despite having found a martingale connected to the binomial model, in Proposition 5.5.6, we won't use martingales to analyse our financial models – yet. That will come in Chapter 15, once we have moved into continuous time and introduced the Black-Scholes model.

7.1 The martingale transform

If M is a stochastic process and C is an adapted process, we define the *martingale transform* of C by M

$$(C \circ M)_n = \sum_{i=1}^{n} C_{i-1}(M_i - M_{i-1}).$$

Here, by convention, we set $(C \circ M)_0 = 0$.

If M is a martingale, the process $(C \circ M)_n$ can be thought of as our winnings after n plays of a game. Here, at round i, a bet of C_{i-1} is made, and the change to our resulting wealth is $C_{i-1}(M_i - M_{i-1})$. For example, if $C_i \equiv 1$ and M_n is the simple random walk $M_n = \sum_1^n X_i$ then $M_i - M_{i-1} = X_i$, so we win/lose each round with even chances; we bet 1 on each round, if we win we get our money back doubled, if we lose we get nothing back.

We place the bet $C_{i-1} \in m\mathcal{F}_{i-1}$ on play i, meaning that we must place our i^{th} bet using only the information we gained during the first $i-1$ plays. In particular, we don't yet know the result of the i^{th} play. So our filtration here is $\mathcal{F}_n = \sigma(C_i, X_i\,;\, i \le n)$.

Theorem 7.1.1 *If M is a martingale and C is adapted and bounded, then $(C \circ M)_n$ is also a martingale.*

Similarly, if M is a supermartingale (resp. submartingale), and C is adapted, bounded and non-negative, then $(C \circ M)_n$ is also a supermartingale martingale (resp. submartingale).

PROOF: Let M be a martingale. Write $Y = C \circ M$. We have $C_n \in m\mathcal{F}_n$ and $X_n \in m\mathcal{F}_n$, so Proposition 2.2.6 implies that $Y_n \in m\mathcal{F}_n$. Since $|Cn| \le c$ for some c and all n, we have

$$\mathbb{E}|Y_n| \le \sum_{k=1}^{n} \mathbb{E}|C_{k-1}(M_k - M_{k-1})| \le c \sum_{k=1}^{n} \mathbb{E}|M_k| + \mathbb{E}|M_{k-1}| < \infty.$$

So $Y_n \in L^1$. Since C_{n-1} is $\mathcal{F}_{n-1}$-measurable, by linearity of conditional expectation, the taking out what is known rule and the martingale property of M, we have

$$\begin{aligned}
\mathbb{E}[Y_n \,|\, \mathcal{F}_{n-1}] &= \mathbb{E}[Y_{n-1} + C_{n-1}(M_n - M_{n-1}) \,|\, \mathcal{F}_{n-1}] \\
&= Y_{n-1} + C_{n-1}\mathbb{E}[M_n - M_{n-1} \,|\, \mathcal{F}_{n-1}] \\
&= Y_{n-1} + C_{n-1}(\mathbb{E}[M_n|\mathcal{F}_{n-1}] - M_{n-1}) \\
&= Y_{n-1}.
\end{aligned}$$

Hence Y is a martingale.

The argument is easily adapted to prove the second statement, e.g. for a supermartingale M, $\mathbb{E}[M_n - M_{n-1} \,|\, \mathcal{F}_{n-1}] \le 0$. Note that in these cases it is important that C is non-negative. ■

7.2 Roulette

The martingale transform is a useful theoretical tool (see e.g. Sections 7.3, 8.2 and 12.1), but it also provides a framework to model casino games. We illustrate this with Roulette.

In roulette, a metal ball lies inside of a spinning wheel. The wheel is divided into 37 segments, of which 18 are black, 18 are red, and 1 is green. The wheel is spun, and the ball spins with it, eventually coming to rest in one of the 37 segments. If the roulette wheel is manufactured properly, the ball lands in each segment with probability $\frac{1}{37}$ and the result of each spin is independent.

On each spin, a player can bet an amount of money C. The player chooses either red or black. If the ball lands on the colour of their choice, they get their bet of C returned and win an additional C. Otherwise, the casino takes the money and the player gets nothing.

The key point is that players can only bet on red or black. If the ball lands on green, the casino takes *everyones* money.

Remark 7.2.1 Thinking more generally, most casino games fit into this mould – there is a very small bias towards the casino earning money. This bias is known as the 'house advantage'.

In each round of roulette, a players probability of winning is $\frac{18}{37}$ (it does not matter which colour they pick). Let (X_n) be a sequence of i.i.d. random variables such that

$$X_n = \begin{cases} 1 & \text{with probability } \frac{18}{37} \\ -1 & \text{with probability } \frac{19}{37} \end{cases}$$

Naturally, the first case corresponds to the player winning game n and the second to losing. We define

$$M_n = \sum_{i=1}^{n} X_n.$$

Then, the value of $M_n - M_{n-1} = X_n$ is 1 if the player wins game n and -1 if they lose. We take our filtration to be generated by (M_n), so $\mathcal{F}_n = \sigma(M_i\,;\, i \leq n)$.

A player cannot see into the future. So the bet they place on game n must be chosen before the game is played, at time $n-1$ – we write this bet as C_{n-1}, and require that it is $\mathcal{F}_{n-1}$ measurable. Hence, C is adapted. The total profit/loss of the player over time is the martingale transform

$$(C \circ M)_n = \sum_{i=1}^{n} C_{i-1}(M_i - M_{i-1}).$$

We'll now show that (M_n) is a supermartingale. We have $M_n \in m\mathcal{F}_n$ and since $|M_n| \leq n$ we also have $M_n \in L^1$. Lastly,

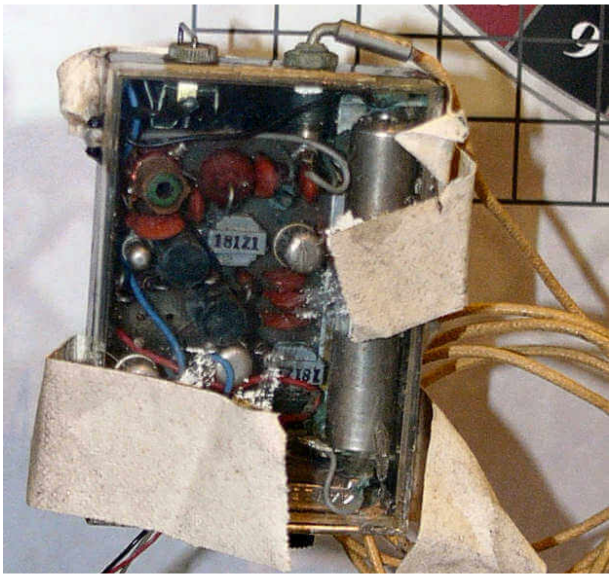
$$\begin{aligned}
\mathbb{E}[M_{n+1}|\mathcal{F}_n] &= \mathbb{E}\left[X_{n+1} + M_n \mid \mathcal{F}_n\right] \\
&= \mathbb{E}[X_{n+1}|\mathcal{F}_n] + M_n \\
&= \mathbb{E}[X_{n+1}] + M_n \\
&\leq M_n.
\end{aligned}$$

Here, the second line follows by linearity and the taking out what is known rule. The third line follows because X_{n+1} is independent of $\mathcal{F}_n$, and the last line follows because $\mathbb{E}[X_{n+1}] = \frac{-1}{37} < 0$.

So, (M_n) is a supermartingale and (C_n) is adapted. Theorem 7.1.1 applies and tells us that $(C \circ M)_n$ is a supermartingale. We'll continue this story in Section 7.4.

Remark 7.2.2 There have been ingenious attempts to win money at Roulette, often through hidden technology or by exploiting mechanical flaws (which can slightly bias the odds), mixed with probability theory.

In 1961, Edward O. Thorp (a professor of mathematics) and Claude Shannon (a professor of electrical engineering) created the worlds first wearable computer, which timed the movements of the ball and wheel, and used this information to try and predict roughly where the ball would land. Here's the machine itself:



Information was input to the computer by its wearer, who silently tapped their foot as the Roulette wheel spun. By combining this information with elements of probability theory, they believed they could consistently 'beat the casino'. They were very successful, and obtained a return of 144% on their bets. Their method is now illegal; at the time it was not, because no-one had even thought it might be possible.

Of course, most gamblers are not so fortunate.

7.3 The martingale convergence theorem

In this section, we are concerned with almost sure convergence of supermartingales (M_n) as $n \to \infty$. Naturally, martingales are a special case and submartingales can be handled through multiplying by -1. We'll need the following definition:

Definition 7.3.1 Let $p \in [1,\infty)$. We say that a stochastic process (X_n) is uniformly bounded in L^p if there exists some $M < \infty$ such that, for all n,

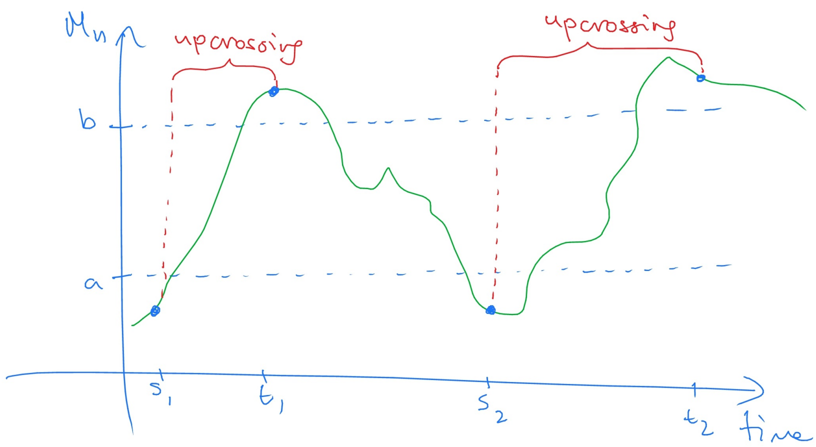
$$\mathbb{E}\left[|X_n|^p\right] \le M.$$

As usual, we'll mostly only be concerned with the cases $p = 1, 2$.

Let (M_n) be a stochastic process and fix $a < b$. We define $U_N[a,b]$ to be the number of *upcrossings* made in the interval $[a,b]$ by $M_1, \dots, M_N$. That is, $U_n[a,b]$ is the largest k such there exists

$$0 \le s_1 < t_2 < \ldots < s_k < t_k \le N \quad \text{such that} \quad M_{s_i} \le a, M_{t_i} > b \ \text{ for all } i = 1, \dots, k.$$

This definition is best understood through a picture:



Note that, for convenience, we draw M_n as a continuous (green line) although in fact it only changing value at discrete times.

Studying upcrossings is key to establishing almost sure convergence of supermartingales. To see why upcrossings are important, note that if $(c_n) \subseteq \mathbb{R}$ is a (deterministic) sequence and $c_n \to c$, for some $c \in \mathbb{R}$, then there is no interval $[a,b]$ $a < b$ such that $(c_n)_{n=1}^{\infty}$ makes infinitely many upcrossings of $[a,b]$; if there was then (c_n) would oscillate and couldn't converge.

Note that $U_N[a,b]$ is an increasing function of N, and define $U_\infty[a,b]$ by

$$U_\infty[a,b](\omega) = \lim_{N\uparrow\infty} U_N[a,b](\omega). \tag{7.1}$$

This is an almost surely limit, since it holds for each $\omega \in \Omega$. With this definition, $U_\infty[a,b]$ could potentially be infinite, but we can prove that it is not.

Lemma 7.3.2 *Let M be a supermartingale. Then*

$$(b-a)\mathbb{E}[U_N[a,b]] \le \mathbb{E}[|M_N - a|].$$

PROOF: Let $C_1 = \mathbb{1}\{M_0 < a\}$ and recursively define

$$C_n = \mathbb{1}\{C_{n-1} = 1, M_n \le b\} + \mathbb{1}\{C_{n-1} = 0, M_n \le a\}.$$

The behaviour of C_n is that, when X enters the region below a, C_n starts taking the value 1. It will continue to take the value 1 until M enters the region above b, at which point C_n will start taking the value 0. It will continue to take the value 0 until M enters the region below a, and so on. Hence,

$$(C \circ M)_N = \sum_{k=1}^{N} C_{k-1}(M_k - M_{k-1}) \ge (b-a)U_N[a,b] - |M_N - a|. \tag{7.2}$$

That is, each upcrossing of $[a,b]$ by M picks up at least $(b-a)$; the final term corresponds to an upcrossing that M might have started but not finished.

Note that C_n is adapted, bounded and non-negative. Hence, by Theorem 7.1.1 we have that $C \circ M$ is a supermartingale. Thus $\mathbb{E}[(C \circ M)_N] \le 0$, which combined with (7.2) proves the given result. ∎

Lemma 7.3.3 *Suppose M is a supermartingale and uniformly bounded in L^1. Then $P[U_\infty[a,b] = \infty] = 0$.*

PROOF: From Lemma 7.3.2 we have

$$(b-a)\mathbb{E}[U_N[a,b]] \le |a| + \sup_{n\in\mathbb{N}} \mathbb{E}|M_n|. \tag{7.3}$$

We have that $U_N[a,b]$ is increasing, as N increases, and the definition of $U_\infty[a,b]$ in (7.1) gives that that $U_N[a,n] \to U_\infty[a,b]$ almost surely as $N \to \infty$. Hence, by the monotone convergence theorem, $\mathbb{E}[U_N[a,b]] \to \mathbb{E}[U_\infty[a,b]]$, and so by letting $N \to \infty$ in (7.3) we have

$$(b-a)\mathbb{E}[U_\infty[a,b]] \le |a| + \sup_{n\in\mathbb{N}} \mathbb{E}|M_n| < \infty,$$

which implies that $\mathbb{P}[U_\infty[a,b] < \infty] = 1$. ∎

Essentially, Lemma 7.3.3 says that the paths of M cannot oscillate indefinitely. This is the crucial ingredient of the martingale convergence theorem.

Theorem 7.3.4 (Martingale Convergence Theorem I) *Suppose M is a supermartingale and uniformly bounded in L^1. Then the limit $M_n \stackrel{a.s.}{\to} M_\infty$ exists and $\mathbb{P}[|M_\infty| < \infty] = 1$.*

PROOF: Define

$$\Lambda_{a,b} = \{\omega : \text{ for infinitely many } n, M_n(\omega) < a\} \cap \{\omega : \text{ for infinitely many } n, M_n(\omega) > b\}.$$

We observe that $\Lambda_{a,b} \subseteq \{U_\infty[a,b] = \infty\}$, which has probability 0 by Lemma 7.3.3. But since

$$\{\omega : M_n(\omega) \text{ does not converge to a limit in } [-\infty,\infty]\} = \bigcup_{\substack{a,b\in\mathbb{Q}\\ a<b}} \Lambda_{a,b},$$

we have that

$$\mathbb{P}[M_n \text{ converges to some } M_\infty \in [-\infty,+\infty]] = 1$$

which proves the first part of the theorem.

To prove the second part we will use an inequality that holds for any convergent sequence $M_n \stackrel{a.s.}{\to} M_\infty$ of random variables:

$$\mathbb{E}[|M_\infty|] \leq \sup_{n\in\mathbb{N}} \mathbb{E}[|M_n|]. \tag{7.4}$$

This inequality can be proved using the monotone convergence theorem and some careful analysis, see exercise **7.14**. Since we assumed that (M_n) is uniformly bounded in L^1, (7.4) gives us that $\mathbb{E}[|M_\infty|] < \infty$. Hence, $\mathbb{P}[|M_\infty| = \infty] = 0$ (or else the expected value would be infinite). ∎

One useful note is that if M_n is a non-negative supermartingale then we have $\mathbb{E}[|M_n|] = \mathbb{E}[M_n] \leq \mathbb{E}[M_0]$, so in this case M is uniformly bounded in L^1.

Theorem 7.3.4 has one big disadvantage: it cannot tell us anything about the limit M_∞, except that it is finite. To gain more information about M_∞, we need an extra condition.

Corollary 7.3.5 (Martingale Convergence Theorem II) *In the setting of Theorem 7.3.4, suppose additionally that (M_n) is uniformly bounded in L^2. Then $M_n \to M_\infty$ in both L^1 and L^2, and*

$$\lim_{n\to\infty} \mathbb{E}[M_n] = \mathbb{E}[M_\infty], \qquad \lim_{n\to\infty} \mathrm{var}(M_n) \to \mathrm{var}(M_\infty).$$

The proof of Corollary 7.3.5 is outside of the scope of our course.

7.4 Long term behaviour of stochastic processes

Our next step is to use the martingale convergence theorem to look at the behaviour as $t \to \infty$ of various stochastic processes. We will begin with Roulette, from Section 7.2, and then move on to the trio of stochastic processes (random walks, urns, branching processes) that we introduced in Chapter 4.

One fact, from real analysis, that we will find useful is the following:

Lemma 7.4.1 *Let (a_n) be an integer valued sequence, and suppose that $a_n \to a$ as $n \to \infty$, where $a \in \mathbb{R}$. Then (a_n) is eventually equal to a constant: there exists $N \in \mathbb{N}$ such that $a_n = a$ for all $n \geq N$.*

PROOF: Put $\epsilon = \frac{1}{3}$ into the definition of a convergent real sequence, and we obtain that there exists $N \in \mathbb{N}$ such that $|a_n - a| \leq \frac{1}{3}$ for all $n \geq N$. Hence, for $n \geq N$ we have

$$\begin{aligned} |a_n - a_{n+1}| = |a_n - a + a - a_{n+1}| &\leq |a_n - a| + |a - a_{n+1}| \\ &\leq \tfrac{1}{3} + \tfrac{1}{3} \\ &= \tfrac{2}{3}. \end{aligned}$$

Since $\frac{2}{3} < 1$, and a_n takes only integer values, this means that $a_n = a_{n+1}$. Since this holds for all $n \geq N$, we must have $a_N = a_{N+1} = a_{N+2} = \ldots$, which means that $a_n \to a_N$ as $N \to \infty$ (and hence $a = a_N$). ∎

Long term behaviour of Roulette

Let us think about what happens to our Roulette player, from Section 7.2.

Recall that our gambler bets an amount $C_{n-1} \in \mathcal{F}_{n-1}$ on play n. On each play, our gambler increases his wealth, by C_{n-1}, with probability $\frac{18}{37}$, or decreases his wealth, by C_{n-1}, with probability $\frac{19}{37}$. We showed in Section 7.2 that his profit/loss after n plays was a supermartingale, $(C \circ M)_n$. Thus, if our gambler starts with an amount of money $K \in \mathbb{N}$ at time 0, then at time n they have an amount of money given by

$$W_n = K + (C \circ M)_n. \tag{7.5}$$

We'll assume that our gambler keeps betting until they run out of money, so we have $C_n \geq 1$ and $W_n \geq 0$ for all n before they run out of money (and, if and when they do run out, $C_n = 0$ for all remaining n). Since money is a discrete quantity, we can assume that K and C_n are integers, hence W_n is also always an integer.

From Section 7.2 we know that $(C \circ M)_n$ is a supermartingale, hence W_n is also a supermartingale. Since $W_n \geq 0$ we have $\mathbb{E}[|W_n|] = \mathbb{E}[W_n] \leq \mathbb{E}[W_0]$ so (W_n) is uniformly bounded in L^1. Hence, by the martingale convergence theorem, W_n is almost surely convergent.

By Lemma 7.4.1, the only way that W_n can converge is if it becomes constant, eventually. Since each play results in a win (an increase) or a loss (a decrease) the only way W_n can become eventually constant is if our gambler has lost all his money i.e. for some (random) $N \in \mathbb{N}$, for all $n \geq N$, we have $W_n = 0$. Thus:

Lemma 7.4.2 *Almost surely, a roulette player will eventually lose all their money.*

Long term behaviour of urn processes

We will look at the Pólya urn process introduced in Section 4.2. Recall that we start our urn with 2 balls, one red and one black. Then, at each time $n = 1, 2, \ldots,$ we draw a ball, uniformly at random, from the urn, and replace it alongside an additional ball of the same colour. Thus, at time n (which means: after the n^{th} draw is completed) the urn contains $n + 2$ balls.

Recall also that we write B_n for the number of red balls in the urn at time n, and

$$M_n = \frac{B_n}{n+2} \tag{7.6}$$

for the fraction of red balls in the urn at time n. We have shown in Section 4.2 that M_n is a martingale. Since $M_n \in [0,1]$, we have $\mathbb{E}[|M_n|] \le 1$, hence (M_n) is uniformly bounded in L^1 and the martingale convergence theorem applies. Therefore, there exists a random variable M_∞ such that $M_n \stackrel{a.s.}{\to} M_\infty$. We will show that:

Proposition 7.4.3 *M_∞ has the uniform distribution on* $[0, 1]$.

Remark 7.4.4 Results like Proposition 7.4.3 are extremely useful, in stochastic modelling. It provides us with the following 'rule of thumb': if n is large, then the fraction of red balls in our urn is approximately a uniform distribution on $[0, 1]$. Therefore, we can approximate a complicated object (our urn) with a much simpler object (a uniform random variable).

We now aim to prove Proposition 7.4.3, which means we must find out the distribution of M_∞. In order to do so, we will use Lemma 6.1.2, which tells us that $M_n \stackrel{d}{\to} M_\infty$, that is

$$\mathbb{P}[M_n \le x] \to \mathbb{P}[M_\infty \le x] \tag{7.7}$$

for all x, except possibly those for which $\mathbb{P}[M_\infty = x] > 0$. We will begin with a surprising lemma that tells us the distribution of B_n. Then, using (7.6), we will be able to find the distribution of M_n, from which (7.7) will then tell us the distribution of M_∞.

Lemma 7.4.5 *For all $k = 1, \ldots, n+1$, it holds that* $\mathbb{P}[B_n = k] = \frac{1}{n+1}$.

PROOF: Let A be the event that the first k balls drawn are red, and the next j balls drawn are black. Then,

$$\mathbb{P}[A] = \frac{1}{2}\frac{2}{3}\ldots\frac{k}{k+1}\frac{1}{k+2}\frac{2}{k+3}\ldots\frac{j}{j+k+1}. \tag{7.8}$$

Here, each fraction corresponds (in order) to the probability of getting a red/black ball (as appropriate) on the corresponding draw, given the results of previous draws. For example, $\frac{1}{2}$ is the probability that the first draw results in a red ball, after which the urn contains 2 red balls and 1 black ball, so $\frac{2}{3}$ is the probability that the second draw results in a red ball, and so on. From (7.8) we have $\mathbb{P}[A] = \frac{j!k!}{(j+k+1)!}$.

Here is the clever part: we note that drawing k red balls and j black balls, in the first $j + k$ draws but in a different order, would have the *same* probability. We would simply obtain the numerators in (7.8) in a different order. There are $\binom{j+k}{k}$ possible different orders (i.e. we must choose k time-points from $j + k$ times at which to pick the red balls). Hence, the probability that we draw k red and j black in the first $j + k$ draws is

$$\binom{j+k}{k}\frac{j!k!}{(j+k+1)!} = \frac{(j+k)!}{k!(j+k-k)!}\frac{j!k!}{(j+k+1)!} = \frac{1}{j+k+1}.$$

We set $j = n - k$, to obtain the probability of drawing (and thus adding) k red balls within the first $j + k = n$ draws. Hence $\mathbb{P}[B_n = k + 1] = \frac{1}{n+1}$ for all $k = 0, 1, \ldots, n$. Setting $k' = k + 1$, we recover the statement of the lemma. ■

Proof: [Of Proposition 7.4.3.] Now that we know the distribution of B_n, we can use (7.6) to find out the distribution of M_n. Lemma 7.4.5 tells us that B_n is uniformly distribution on the set $\{1, \ldots, n+1\}$. Hence M_n is uniform on $\{\frac{1}{n+2}, \frac{2}{n+2}, \ldots, \frac{n+1}{n+2}\}$. This gives us that, for $x \in (0, 1)$,

$$\mathbb{P}[M_n \le x] = \frac{1}{n+1} \times \lfloor x(n+2) \rfloor.$$

Here, $\lfloor x(n+2) \rfloor$, which denotes the integer part of $x(n+2)$, is equal to the number of elements of $\{\frac{1}{n+2}, \frac{2}{n+2}, \ldots, \frac{n+1}{n+2}\}$ that are $\le x$. Since $x(n+2) - 1 \le \lfloor x(n+2) \rfloor \le x(n+2)$ we obtain that

$$\frac{x(n+2) - 1}{n+1} \le \mathbb{P}[M_n \le x] \le \frac{x(n+2)}{n+1}$$

and the sandwich rule tells us that $\lim_{n\to\infty} \mathbb{P}[M_n \le x] = x$. By (7.7), this means that

$$\mathbb{P}[M_\infty \le x] = x \qquad \text{for all } x \in (0, 1).$$

Therefore, M_∞ has a uniform distribution on $(0, 1)$. This proves Proposition 7.4.3. ■

Extensions and applications of urn processes

We could change the rules of our urn process, to create a new kind of urn – for example whenever we draw a red ball we could add two new red balls, rather than just one. Urn processes are typically very ‘sensitive’, meaning that a small change in the process, or its initial conditions, can have a big effect on the long term behaviour. You can investigate some examples of this in exercises **7.7**, **7.8** and **7.9**.

Urn processes are often used in models in which the ‘next step’ of the process depends on sampling from its current state. Recall that in random walks we have $S_{n+1} = S_n + X_{n+1}$, where X_{n+1} is independent of S_n; so for random walks the next step of the process is independent of its current state. By contrast, in Pólya’s urn, whether the next ball we add is red (or black) depends on the current state of the urn; it depends on which colour ball we draw.

For example, we might look at people choosing which restaurant they have dinner at. Consider two restaurants, say *Café Rouge* (which is red) and *Le Chat Noir* (which is black). We think of customers in Café Rouge as red balls, and customers in Le Chat Noir as black balls. A newly arriving customer (i.e. a new ball that we add to the urn) choosing which restaurant to visit is more likely, but not certain, to choose the restaurant which currently has the most customers.

Other examples include the growth of social networks (people tend to befriend people who already have many friends), machine learning (used in techniques for becoming successively more confident about partially learned information), and maximal safe dosage estimation in clinical trials (complicated reasons). Exercise **7.8** features an example of an urn process that is the basis for modern models of evolution (successful organisms tend to have more descendants, who are themselves more likely to inherit genes that will make them successful).

Long term behaviour of Galton-Watson processes

The martingale convergence theorem can tell us about the long term behaviour of stochastic processes. In this section we focus on the Galton-Watson process, which we introduced in Section 4.3.

Let us recall the notation from Section 4.3. Let X_i^n, where $n, i \geq 1$, be i.i.d. nonnegative integer-valued random variables with common distribution G. Define a sequence (Z_n) by $Z_0 = 1$ and

$$Z_{n+1} = \begin{cases} X_1^{n+1} + \ldots + X_{Z_n}^{n+1}, & \text{if } Z_n > 0 \\ 0, & \text{if } Z_n = 0 \end{cases} \tag{7.9}$$

Then (Z_n) is the Galton-Watson process. In Section 4.3 we showed that, writing $\mu = \mathbb{E}[G]$,

$$M_n = \frac{Z_n}{\mu^n}$$

was a martingale with mean $\mathbb{E}[M_n] = 1$.

We now look to describe the long term behaviour of the process Z_n, meaning that we want to know how Z_n behaves as $n \to \infty$. We'll consider three cases: $\mu < 1$, $\mu = 1$, and $\mu > 1$. Before we start, its important to note that, if $Z_N = 0$ for any $N \in \mathbb{N}$, then $Z_n = 0$ for all $n \geq N$. When this happens, it is said that the process *dies out*.

We'll need the martingale convergence theorem. We have that M_n is a martingale, with expected value 1, and since $M_n \geq 0$ we have $\mathbb{E}[|M_n|] = 1$. Hence, (M_n) is uniformly bounded in L^1 and by the martingale convergence theorem (Theorem 7.3.4) we have that the almost sure limit

$$\lim_{n\to\infty} M_n = M_\infty$$

exists. Since $M_n \geq 0$, we have $M_\infty \in [0, \infty)$.

For the rest of this section, we'll assume that the offspring distribution G is not deterministic. Obviously, if G was deterministic and equal to say, c, then we would simply have $Z_n = c^n$ for all n.

Lemma 7.4.6 *Suppose that $\mu = 1$. Then $\mathbb{P}[Z_n$ dies out$] = 1$.*

Proof: In this case $M_n = Z_n$. So, we have $Z_n \to M_\infty$ almost surely. The offspring distribution is not deterministic, so for as long as $Z_n \neq 0$ the value of Z_n will continue to change over time. Each such change in value is of magnitude at least 1, hence by Lemma 7.4.1 the only way Z_n can converge is if Z_n is eventually zero. Therefore, since Z_n does converge, in this case we must have $\mathbb{P}[Z_n \text{ dies out}] = 1$ (and $M_\infty = 0$). ∎

Lemma 7.4.7 *Suppose that $\mu < 1$. Then $\mathbb{P}[Z_n$ dies out$] = 1$.*

Sketch of Proof: Recall our graphical representation of the Galton-Watson process in Section 4.3. Here's the key idea: when $\mu < 1$, on average we expect each individual to have less children than when $\mu = 1$. Therefore, if the Galton-Watson process dies out when $\mu = 1$ (as is the case, from Lemma 7.4.6), it should also die out when $\mu < 1$.

Our idea is not quite a proof, because the size of the Galton-Watson process Z_n is random, and not necessarily equal to its expected size $\mathbb{E}[Z_n] = \mu^n$. To make the idea into a proof we will need to find two Galton-Watson processes, Z_n and $\widetilde{Z}_n$, such that $0 \leq Z_n \leq \widetilde{Z}_n$, with $\mu < 1$ in Z_n

and $\mu = 1$ in $\widetilde{Z}_n$. Then, when $\widetilde{Z}_n$ becomes extinct, so does Z_n. This type of argument is known as a *coupling*, meaning that we use one stochastic process to control another. See exercise **7.15** for how to do it. ∎

Lemma 7.4.8 *Suppose that $\mu > 1$ and $\sigma^2 < \infty$. Then $\mathbb{P}[Z_n \to \infty] > 0$.*

PROOF: The first step of the argument is to show that, for all $n \in \mathbb{N}$,

$$\mathbb{E}[M_{n+1}^2] = \mathbb{E}[M_n^2] + \frac{\sigma^2}{\mu^{n+2}}. \tag{7.10}$$

You can find this calculation as exercise **7.13**. It is similar to the calculation that we did to find $\mathbb{E}[M_n]$ in Section 4.3, but a bit harder because of the square.

By iterating (7.10) and noting that $\mathbb{E}[M_0] = 1$, we obtain that

$$\mathbb{E}[M_{n+1}^2] = 1 + \sum_{i=1}^{n} \frac{\sigma^2}{\mu^{i+2}}.$$

Hence,

$$\mathbb{E}[M_{n+1}^2] \le 1 + \frac{\sigma^2}{\mu^2} \sum_{i=1}^{\infty} \frac{1}{\mu^i},$$

which is finite because $\mu > 1$. Therefore, (M_n) is uniformly bounded in L^2, and the second martingale convergence theorem applies. In particular, this gives us that

$$\mathbb{E}[M_n] \to \mathbb{E}[M_\infty].$$

Noting that $\mathbb{E}[M_n] = 1$ for all n, we obtain that $\mathbb{E}[M_\infty] = 1$. Hence, $\mathbb{P}[M_\infty > 0] > 0$. On the event that $M_\infty > 0$, we have $M_n = \frac{Z_n}{\mu^n} \to M_\infty > 0$. Since $\mu^n \to \infty$, this means that when $M_\infty > 0$ we must also have $Z_n \to \infty$. Hence, $\mathbb{P}[Z_n \to \infty] > 0$. ∎

The behaviour $Z_n \to \infty$ is often called 'explosion', reflecting the fact that (when it happens) each generation tends to contain more and more individuals than the previous one. The case of $\mu > 1$ and $\sigma^2 = \infty$ behaves in the same way as Lemma 7.4.8, but the proof is more difficult and we don't study it in this course.

Extensions and applications of the Galton-Watson process

The Galton-Watson process can be extended in many ways. For example, we might vary the offspring distribution used in each generation, or add a mechanism that allows individuals to produce their children gradually across multiple time steps. The general term for such processes is 'branching processes'. Most branching processes display the same type of long-term behaviour as we have uncovered in this section: if the offspring production is too low (on average) then they are sure to die out, but, if the offspring production is high enough that dying out is not a certainty then (instead) there is a chance of explosion to ∞.

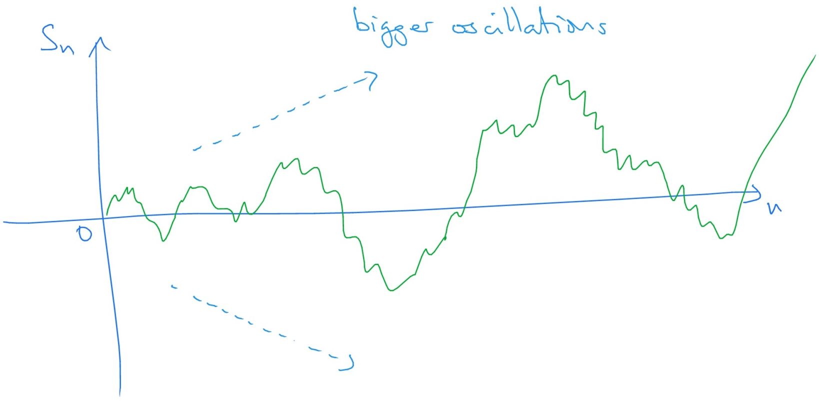
Branching processes are the basis of stochastic models in which existing objects reproduce, or break up, into several new objects. Examples are the spread of contagious disease, crushing rocks, nuclear fission, tumour growth, and so on. In Chapter 19 (which is part of MAS61023 only) we will use branching processes to model contagion of unpaid debts in banking networks.

Branching processes are used by biologists to model population growth, and they are good models for populations that are rapidly increasing in size. (For populations that have reached a roughly constant size, models based on urn processes, such as that of exercise **7.8**, are more effective.)

Long term behaviour of random walks (⊘)

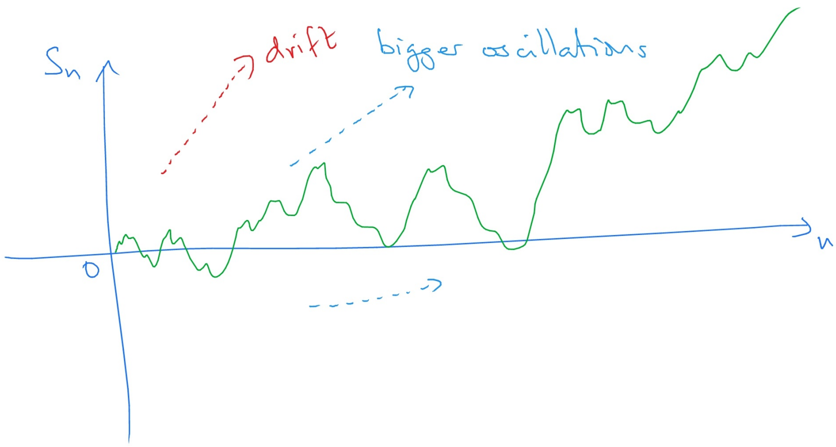
We now consider the long term behaviour of random walks. This section won't include proofs, and for that reason it is marked with a (⊘) and is off syllabus. We will cover this chapter in lectures, because it contain interesting information that provides intuition for (on-syllabus) results in future chapters. For those of you taking MAS61023 a full set of proofs, using some new techniques, are included in your independent reading in Chapters 8 and 9. However, *some parts* of the results discussed below can be proved using the techniques we've already studied; exercises of this type are on-syllabus for everyone, and you'll see some at the end of this chapter.

Firstly, take S_n to be the simple symmetric random walk. It turns out that S_n will oscillate as $n \to \infty$, and that the magnitude of the largest oscillations will tend to infinity as $n \to \infty$ (almost surely). So, our picture of the long-term behaviour of the symmetric random walk looks like:



Note that we draw the picture as a continuous line, for convenience, but in fact the random walk is jumping between integer values.

Next, let S_n be the asymmetric random walk, with $p > q$, so the walk drifts upwards. The asymmetric random walk also experiences oscillations growing larger and larger, but the drift upwards is strong enough that, in fact $S_n \stackrel{a.s.}{\to} \infty$ as $n \to \infty$. So, our picture of the long-term behaviour of the asymmetric random walk looks like:



Of course, if $q < p$ then S_n would drift downwards, and then $S_n \stackrel{a.s.}{\to} -\infty$.

Extensions and applications of random walks

Random walks can be extended in many ways, such as by adding an environment that influences the behaviour of the walk (e.g. reflection upwards from the origin as in exercise **4.5**, slower movements when above 0, etc), or by having random walkers that walk around a graph ($\mathbb{Z}$, in our case above). More complex extensions involve random walks that are influenced by their own history, for example by having a preference to move into sites that they have not yet visited. In addition, many physical systems can be modelled by systems of several random walks, that interact with each other, for example a setup with multiple random walks happening at the same time (but started from different points in space) in which, whenever random walkers meet, both of them suddenly vanish.

Random walks are the basis of essentially all stochastic modelling that involves particles moving around randomly in space. For example, atoms in gases and liquids, animal movement, heat diffusion, conduction of electricity, transportation networks, internet traffic, and so on.

In Chapters 15 and 16 we will use stock price models that are based on 'Brownian motion', which is itself introduced in Chapter 11 and is the continuous analogue of the symmetric random walk (i.e. no discontinuities). We'll come back to the idea of atoms in gases and liquids in Section 11.1, and we'll also briefly look at heat diffusion in Section 11.3.

7.5 Exercises on Chapter 7

On the martingale transform

7.1 Let $S_n = \sum_{i=1}^n X_i$ be the symmetric random walk, from Section 7.4. In each of the following cases, establish the given formula for $(C \circ S)_n$.

(a) If $C_n = 0$, show that $(C \circ S)_n = 0$.

(b) If $C_n = 1$, show that $(C \circ S)_n = S_n$.

(c) If $C_n = S_n$, show that $(C \circ S)_n = \frac{S_n^2}{2} - \frac{n}{2}$

7.2 Let M_n be a stochastic process. Let $\alpha, \beta \in \mathbb{R}$ and let C_n and D_n be adapted stochastic processes. Let $X_n = \alpha C_n + \beta D_n$. Show that

$$(X \circ M)_n = \alpha (C \circ M)_n + \beta (D \circ M)_n$$

for all n.

On long-term behaviour of stochastic processes

7.3 Let (X_n) be a sequence of independent random variables, with distribution

$$\mathbb{P}[X_n = 1] = \mathbb{P}[X_n = -1] = \frac{1}{2n^2}$$

and $\mathbb{P}[X_n = 0] = 1 - \frac{1}{n^2}$. Define

$$S_n = \sum_{i=1}^n X_i$$

where we take $S_0 = 0$.

(a) Show that S_n is a martingale, and deduce that there exists a real-valued random variable S_∞ such that $S_n \stackrel{a.s.}{\to} S_\infty$ as $n \to \infty$.

(b) Show that, almost surely, there exists some $N \in \mathbb{N}$ such that $X_n = 0$ for all $n \geq N$.

7.4 Write simulations of the symmetric & asymmetric random walks (in a language of your choice). Add functionality to draw the random walk as a graph, with time on the horizontal axis and the value of the walk on the vertical axis.

Look at several samples from your simulations, with e.g. 1000 steps of time, and check that they support the claims made in Section 7.4, about the long-term behaviour of random walks.

Modify your simulation to simulate the random walks in Exercise 7.3 and Question 2 of Assignment 3. Check that your graphs support the result that, in both cases, $S_n \stackrel{a.s.}{\to} S_\infty$. From your graphs, do you notice a difference in behaviour between these two cases?

7.5 Let $M_n = S_n - L_n$ be the martingale defined in exercise 4.5. Show that (M_n) is not uniformly bounded in L^1.

7.6 Recall the Galton-Watson process (Z_n) from Section 7.4, and recall that it is parametrized by its offspring distribution G.

(a) Give an example of an offspring distribution G for which $\mathbb{P}[Z_n \text{ dies out}] = 1$.

(b) Give an example of an offspring distribution G for which $\mathbb{P}[Z_n \text{ dies out}] = 0$.

7.7 Consider the following modification of the Pólya urn process. At time $n = 0$, the urn contains one red ball and one black ball. Then, at each time $n = 1, 2, \ldots$, we draw a ball from the urn. We place this ball back into the urn, and add one ball of the *opposite* colour to the urn; so if we drew a red ball, we would add a black ball, and vice versa.

Therefore, at time n (which means: after the n^{th} draw is complete) the urn contains $n + 2$ balls. Let B_n denote the number of red balls in the urn at time n, and let $M_n = \frac{B_n}{n+2}$ denote the fraction of red balls in the urn at time n.

(a) Calculate $\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n]$ and hence show that M_n is not a martingale, with respect to the filtration $\mathcal{F}_n = \sigma(B_1, \ldots, B_n)$.

(b) Write a simulation of the urn (in a language of your choice) and use your simulation to make a conjecture about the value of the almost sure limit of M_n as $n \to \infty$. Does this limit depend on the initial state of the urn?

7.8 Consider an urn that, at time $n = 0$, contains $K \geq 1$ balls, each of which is either black or red. At each time $n = 1, 2, \ldots$, we do the following, in order:

1. Draw a ball X_1 from the urn, and record its colour. Place X_1 back into the urn.
2. Draw a ball X_2 from the urn, and discard it.
3. Place a new ball, with the same colour as X_1, into the urn.

Thus, for all time, the urn contains exactly K balls. We write M_n for the fraction of red balls in the urn, after the n^{th} iteration of the above steps is complete.

(a) Show that M_n is a martingale.

(b) Show that there exists a random variable M_∞ such that $M_n \stackrel{a.s.}{\to} M_\infty$ as $n \to \infty$, and deduce that $\mathbb{P}[M_\infty = 0 \text{ or } M_\infty = 1] = 1$.

This process is known as the discrete time 'Moran model', and is a model for the evolution of a population that contains a fixed number K of individual organisms – represented as K balls. At each time n, an individual X_2 (is chosen and) dies and an individual X_1 (is chosen and) reproduces.

Although this model is a highly simplified version of reality, with careful enough application it turns out to be very useful. For example, it is the basis for current methods of reconstructing genealogical trees from data obtained by genome sequencing.

7.9 Consider the Pólya urn process from Section 7.4. Suppose that we begin our urn, at time $n = 0$, with two red balls and one black ball. Let M_n denote the resulting fraction of red balls in the urn at time n.

(a) Show that M_n does not converge almost surely to 0.

(b) Write a simulation of the Pólya urn process (in a language of your choice) and compare the effect of different initial conditions on M_∞.

7.10 Let (S_n) denote the simple asymmetric random walk, with $q > p$. In Exercise **4.2** we showed that

$$M_n = (q/p)^{S_n}$$

is a martingale.

(a) Show that there exists a real valued random variable M_∞ such that $M_n \stackrel{a.s.}{\to} M_\infty$.

(b) Deduce that $\mathbb{P}[M_\infty = 0] = 1$ and that (M_n) is not uniformly bounded in L^2.

(c) Use (a) and (b) to show that $S_n \stackrel{a.s.}{\to} -\infty$. Explain briefly why this means that $S_n \stackrel{a.s.}{\to} \infty$ for asymmetric random walks with $p > q$.

7.11 Let S_n denote the symmetric random walk, from Section 7.4. Recall that $S_0 = 0$.

(a) Show that S_n is even when n is even, and odd when n is odd.

(b) Define $p_n = \mathbb{P}[S_n = 0]$. Show that $p_{2n} = \binom{2n}{n} 2^{-2n}$ and $p_{2n+1} = 0$.
(Hint: Count the number of ways to return to zero after precisely $2n$ steps.)

(c) Show that $p_{2(n+1)} = (1 - \frac{1}{2(n+1)})p_{2n}$ for all n, and hence show that $p_{2n} \to 0$ as $n \to \infty$.
(Hint: Use the inequality $1 - x \leq e^{-x}$, which holds for all $x \geq 0$.)

7.12 Let S_n denote the symmetric random walk, from Section 7.4. Let $f : \mathbb{N} \to \mathbb{N}$ be a deterministic function.

(a) Show that if $\frac{S_n}{f(n)}$ is a martingale (for $n \geq 1$), then f is constant.

(b) Show that if $\frac{S_n}{f(n)}$ is a supermartingale (for $n \geq 1$), then f is constant.

Challenge questions

7.13 *In this question we establish the formula* (7.10)*, which we used in the proof of Lemma 7.4.8.*

Let (Z_n) be the Galton-Watson process, with the offspring distribution G. Suppose that $\mathbb{E}[G] = \mu$ and $\mathrm{var}(G) = \sigma^2 < \infty$. Set $M_n = \frac{Z_n}{\mu^n}$.

(a) Show that

$$\mathbb{E}[(M_{n+1} - M_n)^2 \,|\, \mathcal{F}_n] = \frac{Z_n \sigma^2}{\mu^{2(n+1)}},$$

(b) Deduce from part (a) and exercise 3.6 that $\mathbb{E}[M_{n+1}^2] = \mathbb{E}[M_n^2] + \frac{\sigma^2}{\mu^{n+2}}$.

7.14 *In this question we prove the inequality* (7.4)*, which we used in the proof of the martingale convergence theorem.*

Let M_n be a sequence of random variables such that $M_n \stackrel{a.s.}{\to} M_\infty$. Define

$$X_n = \inf_{k \geq n} |M_k|.$$

(a) Explain why (X_n) is an increasing sequence and, hence, why there is a random variable X_∞ such that $X_n \stackrel{a.s.}{\to} X_\infty$.

(b) Show that, for all $\epsilon > 0$ and all $n \in \mathbb{N}$ there exists some $n' \geq n$ such that

$$|M_{n'}| - \epsilon \leq X_n \leq |M_n|.$$

(c) Deduce that $X_\infty = |M_\infty|$.

(d) Check that the monotone convergence theorem applies to (X_n).

(e) Deduce that $\mathbb{E}[|M_\infty|] \leq \sup_{n \in \mathbb{N}} \mathbb{E}[|M_n|]$.

7.15 *In this question we give a rigorous proof of Lemma 7.4.7.*

Let Z_n and X_i^{n+1} be as in (7.9) (i.e. Z_n is a Galton-Watson process) and suppose that $\mathbb{E}[X_i^{n+1}] = \mu < 1$.

Let $\alpha \in [0,1]$. For each i, n we define an independent random variable C_i^{n+1}, with the same distribution as C where $\mathbb{P}[C = 1] = \alpha$ and $\mathbb{P}[C = 0] = 1 - \alpha$. We define

$$\widetilde{X}_i^{n+1} = \begin{cases} 0 & \text{if } X_i^{n+1} = 0 \text{ and } C_i^{n+1} = 0 \\ 1 & \text{if } X_i^{n+1} = 0 \text{ and } C_i^{n+1} = 1 \\ X_i^{n+1} & \text{if } X_i^{n+1} \geq 1. \end{cases} \tag{7.11}$$

Define $\widetilde{Z}_n$ by setting $\widetilde{Z}_0 = 1$, and then using (7.9) with $\widetilde{Z}_n$ in place of Z_n and $\widetilde{X}_i^{n+1}$ in place of X_i^{n+1}. Define $f : [0,1] \to \mathbb{R}$ by $f(\alpha) = \mathbb{E}[X_i^{n+1}]$.

(a) Convince yourself that $\widetilde{Z}_n$ is a Galton-Watson process, with offspring distribution given by (7.11).

(b) Explain briefly why $0 \leq Z_n \leq \widetilde{Z}_n$ for all n.

(c) Show that $f(0) < 1$ and $f(1) \geq 1$. Deduce that there exists a value $\alpha \in [0,1]$ such that $f(\alpha) = 1$.

(d) Show that $\mathbb{P}[Z_n \text{ dies out}] = 1$.

Chapter 8

Further theory of stochastic processes (Δ)

In this chapter we develop some further tools for analysing stochastic processes: the dominated convergence theorem, the optional stopping theorem and the strong Markov property. In chapters 8 and 9 we will write

$$\min(s,t) = s \wedge t, \qquad \max(s,t) = s \vee t.$$

This is common notation in the field of stochastic processes.

Note that this whole chapter is marked with a (Δ) – it is for independent study in MAS61023 but it is not part of MAS352.

8.1 The dominated convergence theorem (Δ)

The monotone convergence theorem, from Section 6.2, applies to sequences of increasing random variables. However, most sequences are not so well behaved, and in these cases we need a more powerful theorem. Note that this section is marked with (Δ), meaning that it is for independent study in MAS61023, but it is not included in MAS352.

Theorem 8.1.1 (Dominated Convergence Theorem) *Let X_n, X be random variables such that:*

1. $X_n \stackrel{\mathbb{P}}{\to} X$.
2. *There exists a random variable $Y \in L^1$ such that, for all n, $|X_n| \leq |Y|$ almost surely.*

Then $\mathbb{E}[X_n] \to \mathbb{E}[X]$ as $n \to \infty$.

The random variable Y is often known as the *dominating function* or *dominating random variable*. You can find a proof of the theorem in MAS31002/61022.

Example 8.1.2 Let $Z \sim N(\mu, \sigma^2)$. Let $X \in L^1$ be any random variable and let

$$X_n = X + \frac{Z}{n}.$$

We can think of X_n as a noisy measurement of the random variable X, where the noise term $\frac{Z}{n}$ becomes smaller as $n \to \infty$.

Let us check the first condition of the theorem. Note that $|X_n - X| = \frac{|Z|}{n}$, which tends to zero as $n \to \infty$ because $|Z| < \infty$. Hence $X_n \stackrel{a.s.}{\to} X$ as $n \to \infty$, which by Lemma 6.1.2 implies $X_n \stackrel{\mathbb{P}}{\to} X$.

Let us now check the second condition, with $Y = |X| + |Z|$. Then $\mathbb{E}[Y] = \mathbb{E}[|X|] + \mathbb{E}[|Z|]$, which is finite since $X \in L^1$ and $Z \in L^1$. Hence, $Y \in L^1$. We have $|X_n| \leq |X| + \frac{1}{n}|Z| \leq Y$.

Therefore, we can apply the dominated convergence theorem and deduce that $\mathbb{E}[X_n] \to \mathbb{E}[X]$ as $n \to \infty$. Of course, we can also calculate $\mathbb{E}[X_n] = \mathbb{E}[X] + \frac{1}{n}\mu$, and check that the result really is true.

In the above example, we could calculate $\mathbb{E}[X_n] = \mathbb{E}[X] + \frac{\mu}{n}$ and notice that $\mathbb{E}[X_n] \to \mathbb{E}[X]$, without using any theorems at all. The real power of the dominated convergence theorem is in situations where we don't specify, or can't easily calculate, the means of X_n or X. See, for example, exercises **8.1** and **8.2**, or the applications in Section 8.2.

If our sequence of random variables (X_n) has $|X_n| \leq c$ for some deterministic constant c, then the dominating function can be taken to be (the deterministic random variable) c. This case is the a very common application, see e.g. exercise **8.1**.

Remark 8.1.3 ($\oslash$) The dominated convergence theorem holds for conditional expectation too; that is with $\mathbb{E}[\cdot]$ replaced by $\mathbb{E}[\cdot \,|\, \mathcal{G}]$. We won't need this result as part of our course.

8.2 The optional stopping theorem (Δ)

Adaptedness, from Definition 3.3.2, captures the idea that we cannot see into the future. This leaves us with the question: as time passes, in what circumstances can we tell if an event has already occurred?

Definition 8.2.1 A map $T : \Omega \to \{0, 1, 2, \ldots, \infty\}$ is called a $(\mathcal{F}_n)$ stopping time if, for all n, the event $\{T \le n\}$ is $\mathcal{F}_n$ measurable.

A stopping time is a random time with the property that, if we have only information from $\mathcal{F}_n$ accessible to us at time n, we are able to decide at any n whether or not T has already happened. It is straightforward to check that T is a stopping time if and only if $\{T = n\}$ is $\mathcal{F}_n$ measurable for all n. This is left for you as exercise **8.4**.

Example 8.2.2 Let $S_n = \sum_{i=1}^n X_i$ be the simple symmetric random walk, with $\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = -1] = \frac{1}{2}$. Let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Then, for any $a \in \mathbb{N}$, the time

$$T = \inf\{n \ge 0\,;\, S_n = a\},$$

which is the first time S_n takes the value a, is a stopping time. It is commonly called the 'hitting time' of a. To see that T is a stopping time we note that

$$\{T \le n\} = \{\text{for some } i \le n \text{ we have } S_i = a\} = \bigcup_{i=0}^{n} \{S_i = a\} = \bigcup_{i=0}^{n} S_i^{-1}(a)$$

Since S_i is $\mathcal{F}_n$ measurable for all $i \le n$, the above equation shows that $\{T \le n\} \in \mathcal{F}_n$.

You will have seen hitting times before – for most of you, in MAS2003 in the context of Markov chains. You might also recall that the random walk S_n can be represented as Markov chains on $\mathbb{Z}$. You will probably have looked at a method for calculating expected hitting times by solving systems of linear equations. This approach is feasible if the state space of a Markov chain is finite; however $\mathbb{Z}$ is infinite. In this section we look at an alternative method, based on martingales.

Recall the notation $\wedge$ and $\vee$ that we introduced at the start of this chapter. That is, $\min(s, t) = s \wedge t$ and $\max(s, t) = s \vee t$.

Lemma 8.2.3 *Let S and T be stopping times with respect to the filtration $(\mathcal{F}_n)$. Then $S \wedge T$ is also a $(\mathcal{F}_n)$ stopping time.*

Proof: Note that

$$\{S \wedge T \le n\} = \{S \le n\} \cup \{T \le n\}.$$

Since S and T are stopping times, both the sets on the right hand side of the above are events in the σ-field $\mathcal{F}_n$. Hence, the event on the left hand side is also in $\mathcal{F}_n$. ■

If T is a stopping time and M is a stochastic process, we define M^T to be the process

$$M_n^T = M_{n \wedge T}.$$

Here $a \wedge b$ denotes the minimum of a and b. To be precise, this means that $M_n^T(\omega) = M_{n \wedge T(\omega)}(\omega)$ for all $\omega \in \Omega$. In Example 8.2.2, S^T would be the random walk S which is stopped (i.e. never moves again) when (if!) it reaches state a.

Lemma 8.2.4 *Let M_n be a martingale (resp. supmartingale, supermartingale) and let T be a stopping time. Then M^T is also a martingale (resp. supmartingale, supermartingale).*

PROOF: Let $C_n := \mathbb{1}\{T-1 \geq n\}$. Note that $\{T-1 \geq n\} = \Omega \setminus \{T \leq n-1\}$, so $\{T \geq n\} \in \mathcal{F}_n$. By Lemma 2.4.2 $C_n \in m\mathcal{F}_{n-1}$. That is, (C_n) is an adapted process. Moreover,

$$(C \circ M)_n = \sum_{k=1}^{n} \mathbb{1}_{k-1 \leq T-1}(M_k - M_{k-1}) = \sum_{k=1}^{n} \mathbb{1}_{k \leq T}(M_k - M_{k-1}) = \sum_{k=1}^{n \wedge T}(M_k - M_{k-1}) = M_{T \wedge n} - M_0.$$

The last equality holds because the sum is telescoping (i.e. the middle terms all cancel each other out). Hence, by Theorem 7.1.1, if M is a martingale (resp. submartingale, supermartingale), $C \circ M$ is also a martingale (resp. submartingale, supermartingale). ■

Theorem 8.2.5 (Doob's Optional Stopping Theorem) *Let M be martingale (resp. submartingale, supermartingale) and let T be a stopping time. Then*

$$\mathbb{E}[M_T] = \mathbb{E}[M_0]$$

(resp. $\geq$, $\leq$) if any one of the following conditions hold:

(a) T is bounded.

(b) $\mathbb{P}[T < \infty] = 1$ and there exists $c \in \mathbb{R}$ such that $|M_n| \leq c$ for all n.

(c) $\mathbb{E}[T] < \infty$ and there exists $c \in \mathbb{R}$ such that $|M_n - M_{n-1}| \leq c$ for all n.

PROOF: We'll prove this for the supermartingale case. The submartingale case then follows by considering $-M$, and the martingale case follows since martingales are both supermartingales and submartingales.

Note that

$$\mathbb{E}[M_{n \wedge T} - M_0] \leq 0, \tag{8.1}$$

because M^T is a supermartingale, by Lemma 8.2.4. For (a), we take $n = \sup_\omega T(\omega)$ and the conclusion follows.

For (b), we use the dominated convergence theorem to let $n \to \infty$ in (8.1). As $n \to \infty$, almost surely $n \wedge T(\omega)$ is eventually equal to $T(\omega)$ (because $\mathbb{P}[T < \infty] = 1$), so $M_{n \wedge T} \to M_T$ almost surely. Since M is bounded, $M_{n \wedge T}$ and M_T are also bounded. So $\mathbb{E}[M_{n \wedge T}] \to \mathbb{E}[M_T]$ and taking limits in (8.1) obtains $\mathbb{E}[M_T - M_0] \leq 0$, which in turn implies that $\mathbb{E}[M_T] \leq \mathbb{E}[M_0]$.

For (c), we will also use the dominated convergence theorem to let $n \to \infty$ in (8.1), but now we need a different way to check its conditions. We observe that

$$|M_{n \wedge T} - M_0| = \left| \sum_{k=1}^{n \wedge T}(M_k - M_{k-1}) \right| \leq T \sup_{n \in \mathbb{N}} |M_n - M_{n-1}|.$$

Since $\mathbb{E}[T(\sup_n |M_n - M_{n-1}|)] \leq c\mathbb{E}[T] < \infty$, we can use the Dominated Convergence Theorem to let $n \to \infty$, and the results follows as in (b). ■

These three sets of conditions (a)-(c) for the optional stopping theorem are listed on the formula sheet, in Appendix B. Sometimes *none* of them apply! See Remark 9.3.5 and the lemma above it for a warning example.

8.3 The stopped σ-field (Δ)

Given a filtration $(\mathcal{F}_n)$ and a stopping time T, it is natural to think about the information available up to and including time T. The following definition makes this concept precise.

Definition 8.3.1 Let $(\mathcal{F}_n)$ be a filtration and let T be a stopping time. Define

$$\mathcal{F}_T = \{A \in \mathcal{F}\,;\, A \cap \{T \le n\} \in \mathcal{F}_n \text{ for all } n\}.$$

In words, $\mathcal{F}_T$ is often known as the σ-field obtained by stopping $(\mathcal{F}_n)$ at time T. If $(\mathcal{F}_n)$ is the generated filtration of (X_n), then $\mathcal{F}_T$ holds the information obtained from $(X_1, X_2, \dots, X_T)$. We will often need to use our intuition when dealing with this concept, but let us formally prove that $\mathcal{F}_T$ is a σ-field.

Lemma 8.3.2 *Let $(\mathcal{F}_n)$ be a filtration and let T be a stopping time. Then $\mathcal{F}_T$ is a σ-field,*

PROOF: We must check the three conditions of Definition 2.1.1. Firstly, for all $n \in \mathbb{N}$ we have $\emptyset \cap \{T \le n\} = \emptyset \in \mathcal{F}_n$, so $\emptyset \in \mathcal{F}_T$. Secondly, if $A \in \mathcal{F}_T$ then $A \cap \{T \le n\} \in \mathcal{F}_n$, and $\{T \le n\} \in \mathcal{F}_n$ because T is a stopping time. Hence

$$(\Omega \setminus A) \cap \{T \le n\} = \{T \le n\} \setminus (A \cap \{T \le n\}) \quad \in \mathcal{F}_n$$

so $\Omega \setminus A \in \mathcal{F}_T$. Thirdly, if $A_m \in \mathcal{F}_T$ for all m, then $A_m \cap \{T \le n\} \in \mathcal{F}_n$ for all m, n. Hence

$$\left(\bigcup_{m=1}^{\infty} A_m\right) \cap \{T \le n\} = \bigcup_{m=1}^{\infty} A_m \cap \{T \le n\} \quad \in \mathcal{F}_n,$$

so $\cup_{m=1}^{\infty} A_m \in \mathcal{F}_T$. ■

8.4 The strong Markov property (Δ)

Let us briefly consider a general filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$, equipped with an adapted stochastic process (X_m). The *Markov property* states that, for any $m \in \mathbb{N}$, the following two processes have the same distribution:

1. the distribution of $(X_m, X_{m+1}, X_{m+2}, \ldots)$ given the information $\sigma(X_m)$;

2. the distribution of $(X_m, X_{m+1}, X_{m+2}, \ldots)$ given the information $\mathcal{F}_m$.

The key point is this: suppose that we are interested in the distribution of $(X_m, X_{m+1}, X_{m+2}, \ldots)$. This might depend on the value of X_m, but if we know the value of X_m then values of $X_1, \ldots, X_{m-1}$ (alongside any other information in $\mathcal{F}_m$) provide us with *no* extra information. Heuristically, the future of (X_m) might depend on its current value, but if we know that current value then we can ignore the past.

The Markov property is very natural. We might even argue that reality itself, with its generated σ-field, has the Markov property, but this is a philosophical question best discussed elsewhere. Most stochastic processes that are used in modelling are Markov processes. It is also natural to replace the time m by a stopping time T. The *strong Markov property* states that $(X_T, X_{T+1}, X_{T+2}, \ldots)$ has the same distribution given the information $\mathcal{F}_T$ as it would do if all we knew was $\sigma(X_T)$.

We will need the strong Markov property for random walks within Chapter 9. However, for random walks we can say something even stronger: after a stopping time time T, a random walk essentially restarts from its current location, and from then on behaves just like the same random walk, with its future movements being independent of its past.

We need some notation to state this precisely. Let $S_n = \sum_{i=1}^n X_i$ where the (X_i) are independent identically distribution random variables; this covers all the examples of random walks that we will study in Chapter 9.

Theorem 8.4.1 *Let T be a stopping time. The following two processes have the same distribution:*

1. *the distribution of $(S_T, S_{T+1}, S_{T+2}, \ldots)$ given the information $\sigma(S_T)$;*

2. *the distribution of $(S_T, S_{T+1}, S_{T+2}, \ldots)$ given the information $\mathcal{F}_T$.*

Moreover, in either case, the distribution of $(S_T, S_{T+1}, S_{T+2}, \ldots)$ is that of a random walk begun at S_T and incremented by an independent copy of X_i at each step of time.

More formally, Theorem 8.4.1 states that for all $n \in \mathbb{N}$ and any function $f : \mathbb{R} \to \mathbb{R}$ we have

$$\mathbb{E}[f(S_{T+n}) \,|\, \mathcal{F}_T] = \mathbb{E}[f(S_{T+n}) \,|\, \sigma(S_T)] = \mathbb{E}[f(S_T + S'_n)]$$

where (S'_n) is an independent copy of (S_n), with identical distribution. Strictly, this equation only holds provided that $f(\cdot) \in L^1$ in each case, but we will always take f to be a nice enough function that this is not a problem.

It is often more useful to apply the strong Markov property in words, and this is common practice in probability theory. We'll do it that way in Chapter 9. We won't include a proof of Theorem 8.4.1 in this course. It isn't hard to prove, but it is a bit tedious, and hopefully you can believe the result without needing a proof.

8.5 Kolmogorov's 0-1 law (⊘)

This section is off-syllabus and is marked with a (⊘). It contains an intriguing fact about sequences of σ-fields, but it lives somewhere in between the material covered within our own course and MAS31002/61022 (Probability with Measure). It is mainly of interest to those taking MAS31002/61022 alongside this course, and we will not use it within our course, so it is best placed off-syllabus. It has a close connection to the second Borel-Cantelli lemma, which is introduced in MAS31002/61022.

Let $(\mathcal{F}_n)_{n\in\mathbb{N}}$ be a sequence of σ-fields, on the same probability space $(\Omega,\mathcal{F},\mathbb{P})$. The *tail σ-field* $\mathcal{T}$ of $(\mathcal{F}_n)$ is defined by

$$\mathcal{T}=\bigcap_{n=1}^{\infty}\sigma(\mathcal{F}_n,\mathcal{F}_{n+1},\ldots). \tag{8.2}$$

Note that $\mathcal{T}$ is a σ-field by Lemma 2.1.5. The intuition here is that σ-field $\mathcal{T}$ contains events that depend only on information 'in the tail' of the sequence $(\mathcal{F}_n)$. That is, if we have an event $E\in\mathcal{T}$, then for any $N\in\mathbb{N}$ we could tell whether E occurred by looking *only* at the occurrence of events $E'\in\mathcal{F}_n$ for $n\geq N$. For example, if (X_n) is a sequence of random variables and $\mathcal{F}_n=\sigma(X_n)$ then the almost sure limit $X_n\stackrel{a.s.}{\to}X$, if it exists, will satisfy $X\in m\mathcal{T}$.

The next result may appear surprising at first. The key point is that the $\mathcal{F}_n$ are assumed to be independent, which means that they have no information in common. Consequently (8.2) implies that $\mathcal{T}$ contains no information.

Theorem 8.5.1 (Kolmogorov's 0-1 law) *Let $(\mathcal{F}_n)$ be a sequence of independent σ-fields and let $\mathcal{T}$ be the associated tail σ-field. If $A\in\mathcal{T}$ then $\mathbb{P}[A]=0$ or $\mathbb{P}[A]=1$.*

PROOF: Let $A\in\mathcal{T}$. Then $A\in\sigma(\mathcal{F}_{n+1},\mathcal{F}_{n+2},\ldots)$ for all $n\in\mathbb{N}$. This means that A is independent of $\sigma(\mathcal{F}_1,\ldots,\mathcal{F}_n)$, for all n. It follows that A is independent of $\sigma(\mathcal{F}_n\,;\,n\in\mathbb{N})$. However, from (8.2) we have that $\mathcal{T}\subseteq\sigma(\mathcal{F}_n\,;\,n\in\mathbb{N})$, which means that A is independent of $\mathcal{T}$. Hence A is independent of A (this is not a typo!), which means that $\mathbb{P}[A]=\mathbb{P}[A\cap A]=\mathbb{P}[A]\mathbb{P}[A]=\mathbb{P}[A]^2$. The only solutions of the equation $x^2=x$ are 0 and 1, hence $\mathbb{P}[A]=0$ or $\mathbb{P}[A]=1$. ■

Remark 8.5.2 Let us assume the same independence assumption as Theorem 8.5.1, and suppose that $X\in m\mathcal{T}$. Then $\{X\leq x\}\in\mathcal{T}$ for all $x\in\mathbb{R}$, so Theorem 8.5.1 gives that $\mathbb{P}[X\leq x]$ is either 0 or 1 for each $x\in\mathbb{R}$. A bit of analysis shows that if we set $c=\inf\{x\in\mathbb{R}\,;\,\mathbb{P}[X\leq x]=1\}$ then in fact $\mathbb{P}[X=c]=1$. Therefore, any random variable that is $\mathcal{T}$ measurable is almost surely equal to a constant.

8.6 Exercises on Chapter 8 (Δ)

On the dominated convergence theorem

8.1 Let U be a random variable that takes values in $(1,\infty)$. Define $X_n = U^{-n}$. Show that $\mathbb{E}[X_n] \to 0$.

8.2 Let X be a random variable in L^1 and set

$$X_n = X\mathbb{1}_{\{|X|\le n\}} = \begin{cases} X & \text{if } |X| \le n, \\ 0 & \text{otherwise.} \end{cases}$$

Show that $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

8.3 Let Z be a random variable taking values in $[1,\infty)$ and for $n \in \mathbb{N}$ define

$$X_n = Z\mathbb{1}_{\{Z\in[n,n+1)\}} = \begin{cases} Z & \text{if } Z \in [n,n+1) \\ 0 & \text{otherwise.} \end{cases} \tag{8.3}$$

(a) Suppose that $Z \in L^1$. Use the dominated convergence theorem to show that $\mathbb{E}[X_n] \to 0$ as $n \to \infty$.

(b) Suppose, instead, that Z is a continuous random variable with probability density function

$$f(x) = \begin{cases} x^{-2} & \text{if } x \ge 1, \\ 0 & \text{otherwise.} \end{cases}$$

and define X_n using (8.3). Show that Z is not in L^1, but that $\mathbb{E}[X_n] \to 0$.

(c) Comment on what part (b) tells us about the dominated convergence theorem.

On stopping times and optional stopping

8.4 Let $(\mathcal{F}_n)$ be a filtration. Check that T is a stopping time if and only if $\{T = n\}$ is $\mathcal{F}_n$ measurable for all n.

8.5 Let $(\mathcal{F}_n)$ be a filtration and let (S_n) be an adapted process, with values in $\mathbb{Z}$ and initial value $S_0 = 0$. Show that $R = \inf\{n \ge 1\,;\, S_n = 0\}$ is a stopping time.

Note: R is known as the first return time of (S_n) to 0.

8.6 Let S and T be stopping times with respect to the same filtration. Show that $S + T$ is also a stopping time.

8.7 Let $(\mathcal{F}_n)$ be a filtration. Let S and T be stopping times with $S \le T$. Show that $\mathcal{F}_S \subseteq \mathcal{F}_T$.

8.8 Recall the Pólya urn process from Section 4.2. Recall that B_n is the number of black balls in the urn at time n, and that $M_n = \frac{B_n}{n+2}$ is the fraction of black balls in the urn at time n. Let T be the first time at which a black ball is drawn from the urn.

(a) Show that $\mathbb{P}[T \ge n] = \frac{1}{n}$.

(b) Use the optional stopping theorem to show that $\mathbb{E}[M_T] = \frac{1}{2}$ and $\mathbb{E}[\frac{1}{T+2}] = \frac{1}{4}$.

8.9 *This question is a continuation of exercise* **7.8** *– recall the urn process from that exercise.*

Let T be the first time at which the urn contains only balls of one colour, and suppose that initially the urn contains r red balls and b black balls. Recall that M_n denotes the fraction of red balls in the urn at time n.

(a) Deduce that $\mathbb{P}[T < \infty] = 1$, and that $\mathbb{P}[M_T = 0 \text{ or } M_T = 1] = 1$.

(b) Show that $\mathbb{E}[M_T] = \frac{r}{r+b}$ and hence deduce that $\mathbb{P}[M_T = 1] = \frac{r}{r+b}$.

8.10 Let $m \in \mathbb{N}$ and $m \geq 2$. At time $n = 0$, an urn contains $2m$ balls, of which m are red and m are blue. At each time $n = 1, \ldots, 2m$ we draw a single ball from the urn; we discard this ball and do not replace it. We continue until the urn is empty. Therefore, at time $n \in \{0, \ldots, 2m\}$ the urn contains $2m - n$ balls.

Let N_n denote the number of red balls remaining in the urn at time n. For $n = 0, \ldots, 2m-1$ let

$$P_n = \frac{N_n}{2m - n}$$

be the fraction of red balls remaining after time n.

(a) Show that P_n is a martingale, with respect to a natural filtration that you should specify.

(b) **[Challenge question]** Let T be the first time at which we draw a red ball. Note that a $(T+1)^{st}$ ball will be drawn, because the urn initially contains at least two red balls. Show that the probability that the $(T+1)^{st}$ ball is red is $\frac{1}{2}$.

Chapter 9

Simple random walks (Δ)

In Section 7.4 we studied the long-term behaviour of the Polya urn and the Galton-Watson process, by taking a limit as the time parameter tended to infinity of a suitably chosen martingale. The following result shows that these techniques do not (or at least, not directly) explain how random walks might behave. In this chapter we will focus on the case of *simple* random walks, which means that they move precisely one unit of space, upwards or downwards, in each step of time.

Lemma 9.0.1 *Let $S_n = \sum_{i=1}^n X_i$ where $\mathbb{P}[X_i = 1] = p$ and $\mathbb{P}[X_i = -1] = 1 - p$, and the (X_i) are independent of each other. Then (S_n) does not converge almost surely to a real valued random variable and (S_n) is not uniformly bounded in L^1.*

PROOF: The process (S_n) is integer valued. From Lemma 7.4.1 we have that, if (S_n) were to converge almost surely, it would have to eventually become constant. By definition we have $|S_{n+1} - S_n| = 1$, so this cannot happen, hence (S_n) does not convergence almost surely.

In the case $p = \frac{1}{2}$ we have that (S_n) is a martingale. If (S_n) was also uniformly bounded in L^1 then the martingale convergence theorem would imply almost sure convergence, so from what we have already proved (S_n) is not uniformly bounded in L^1. For the case $p \neq \frac{1}{2}$, we have shown in (4.3) that $M_n = S_n - n(p-q)$ is a martingale, hence $\mathbb{E}[S_n] = n(p-q)$ and therefore $|\mathbb{E}[S_n]| \to \infty$. By the absolute values property of expectations $|E[S_n]| \leq \mathbb{E}[|S_n|]$, hence $\mathbb{E}[|S_n|] \to \infty$, so (S_n) is not uniformly bounded in L^1. ■

In fact, as we mentioned briefly (and without proof) in Section 7.4, random walks exhibit a mixture of oscillatory behaviour and divergence to $\pm\infty$. We will study this behaviour here, as well as considering some other closely related properties of random walks. The techniques we will use are primarily those developed in Chapter 8.

This chapter will make use of some of the exercises from earlier chapters, including within the proofs. Make sure that you study these too, when they appear, if you have not done so already. As a reminder: the solutions to all of the end-of-chapter exercises can be found within the online version of the lecture notes.

9.1 Exit probabilities (Δ)

We recall the asymmetric random walk from Section 4.1. Let $(X_i)_{i=1}^{\infty}$ be a sequence of i.i.d. random variables. Let $p+q=1$ with $p,q\in(0,1)$, $p\neq q$ and suppose that

$$\mathbb{P}[X_i=1]=p,\qquad \mathbb{P}[X_i=-1]=q.$$

The asymmetric random walk is the stochastic process

$$S_n=\sum_{i=1}^{n}X_i.$$

Set $\mathcal{F}_n=\sigma(X_1,\dots,X_n)$ and note that $(\mathcal{F}_n)$ is the natural filtration of S_n. Recall that we showed in Section 4.1 that the stochastic process $S_n-n(p-q)$ was a martingale. In Exercise **4.2** we found another martingale associated to (S_n), in particular we showed that

$$M_n=(q/p)^{S_n}$$

is a martingale.

Our plan is to use the optional stopping theorem, applied to the martingale (M_n), to obtain information about the hitting times of the asymmetric random walk. Let $T_a=\inf\{n:S_n=a\}$ and $T=T_a\wedge T_b$ for integer $a<0<b$. We aim to calculate $\mathbb{P}[T=T_a]$ and $\mathbb{P}[T=T_b]$. We can show that T_a is a stopping time by noting that

$$\{T_a\leq n\}=\bigcup_{i=0}^{n}\{S_i\leq a\}.$$

Similarly, T_b is a stopping time and it follows from Lemma 8.2.3 that T is also a stopping time.

We now look to apply the optional stopping theorem, using the (b) conditions. For this, we'll need the following lemma.

Lemma 9.1.1 *It holds that* $\mathbb{E}[T]<\infty$.

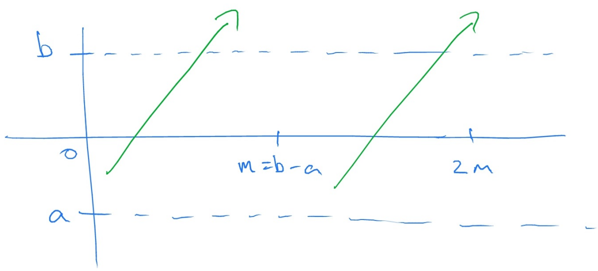
PROOF: Let $m=b-a$. Divide up the random variables (X_n) into sets $A_1,A_2,\dots$ as follows:

$$\overbrace{X_1,X_2,X_3,\dots,X_m}^{A_1},\;\overbrace{X_{m+1},X_{m+2},\dots,X_{2m}}^{A_2},\;\overbrace{X_{2m+1},X_{2m+2},\dots,X_{3m}}^{A_3},\;\overbrace{X_{3m+1},\dots}^{\dots}. \tag{9.1}$$

So as each A_k contains precisely m of the X_is.

Let $E_k=\{\text{for all } X\in A_k,\ X=1\}$ be the event that all random variables in A_k are equal to one. Note that if the event E_k occurs then our random walk moves up by m steps, during time $(k-1)m,\dots,km$.

If T has not happened before time $(k-1)m$ then $a<S_{(k-1)m}<b$. Then, if the event E_k then occurs, we will have $S_{km}\geq a$ and hence $T\leq km$. This is best illustrated with a picture:



m steps up $\Rightarrow$ hit b

We can think of the sequence of events $E_1, E_2, \ldots$ as one chance after another that our random walker has to exit $[a, b]$.

By independence, $\mathbb{P}[E_k] = p^m$. Hence, the random variable $K = \inf\{k \in \mathbb{N}\,;\, E_k \text{ occurs}\}$ is a geometric random variable with success parameter p^m. This means that $K < \infty$ almost surely and that $\mathbb{E}[K] = p^{-m} < \infty$. By definition of K, the event E_K occurs so $T \leq Km$ and by monotonicity of $\mathbb{E}$ we have $\mathbb{E}[T] \leq m\mathbb{E}[K] < \infty$. ∎

From above, we have that M is a martingale. Hence by Lemma 8.2.4, M^T is also a martingale. By definition of T,

$$\begin{aligned} &\text{if } q > p \text{ then } (q/p)^a \leq M_n^T \leq (q/p)^b \\ &\text{if } p > q \text{ then } (q/p)^a \geq M_n^T \geq (q/p)^b \end{aligned}$$

for all n, hence M^T is a bounded martingale. Lemma 9.1.1 implies that $\mathbb{P}[T < \infty] = 1$, so we have that condition (b) of the optional stopping theorem holds for the martingale M^T and the stopping time T. Therefore,

$$\mathbb{E}[M_T^T] = \mathbb{E}[M_0^T]$$

but $M_T^T = M_{T\wedge T} = M_T$ and $M_0^T = M_{0\wedge T} = M_0 = 1$. So we have

$$\mathbb{E}[M_T] = 1.$$

Our next aim is to calculate the probabilities $\mathbb{P}[T = T_a]$ and $\mathbb{P}[T = T_b]$. That is, we want to know which of the two boundaries a and b we actually hit at time T (for example, $\{T = T_a\} = \{S_T = a\}$ is the event that we hit a at time T).

Since $\mathbb{P}[T < \infty] = 1$, we must hit one or other boundary, so we have that

$$\mathbb{P}[T = T_a] + \mathbb{P}[T = T_b] = 1. \tag{9.2}$$

By partitioning the expectation $\mathbb{E}[M_T]$ on whether $\{T = T_a\}$, we have

$$\begin{aligned} 1 &= \mathbb{E}[M_T] && (9.3) \\ &= \mathbb{E}[M_T \mathbb{1}\{T = T_a\}] + \mathbb{E}[M_T \mathbb{1}\{T = T_b\}] \\ &= \mathbb{E}\left[\left(\frac{q}{p}\right)^a \mathbb{1}\{T = T_a\}\right] + \mathbb{E}\left[\left(\frac{q}{p}\right)^b \mathbb{1}\{T = T_b\}\right] \\ &= \mathbb{P}[T = T_a]\left(\frac{q}{p}\right)^a + \mathbb{P}[T = T_b]\left(\frac{q}{p}\right)^b. && (9.4) \end{aligned}$$

Solving the linear equations (9.2) and (9.4), recalling that $p \neq q$, gives that

$$\mathbb{P}[T = T_a] = \frac{(q/p)^b - 1}{(q/p)^b - (q/p)^a}. \tag{9.5}$$

and therefore also

$$\mathbb{P}[T_b = T] = 1 - \mathbb{P}[T = T_a] = \frac{1 - (q/p)^a}{(q/p)^b - (q/p)^a}. \tag{9.6}$$

Note that this formula does not make sense if $p = q = \frac{1}{2}$. In that case our two linear equations above are in fact the same equation, so we cannot solve them to find $\mathbb{P}[T = T_a]$ and $\mathbb{P}[T = T_b]$. However, all is not lost: the case of the symmetric random walk can be handled in similar style to above, but we need a different martingale in place of (M_n). This is left for you to do, with some hints along the way, in Exercise **9.2**

9.2 Stirling's Approximation (Δ)

Stirling's approximation is a fundamental inequality that is used across many different parts of mathematics. It connects the factorial function with the constants π and e, and is often written informally as $n! \approx \sqrt{2\pi n}\,\left(\frac{n}{e}\right)^n$. We mention it here because the distribution of S_n, where (S_n) is a random walk, can sometimes be expressed using formulae involving binomial coefficients. Stirling's approximation is helpful to calculate the limits of such quantities as $n \to \infty$.

It is also helpful to introduce some notation from analysis that you may not have seen before: if (a_n) and (b_n) are real valued sequences then we write

$$a_n \sim b_n \qquad \text{to mean that} \qquad \lim_{n\to\infty} \frac{a_n}{b_n} = 1.$$

We will use this notation throughout the remainder of Chapter 9.

Theorem 9.2.1 *For all $n \in \mathbb{N}$ it holds that*

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{1/(12n+1)} \;<\; n! \;<\; \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{1/(12n)}.$$

This is a result from real analysis so we won't include a proof in this course. Using that $e^{1/n} \to 1$, it is straightforward to check that Theorem 9.2.1 implies the weaker result

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n. \tag{9.7}$$

We will apply Stirling's approximation in this form. Equation (9.7) appears on the formula sheet.

9.3 Long term behaviour: symmetric case (Δ)

In this section we study the simple symmetric random walk. Our goal is to show that (S_n) not only oscillates as $n \to \infty$, but in fact (S_n) visits every $k \in Z$ infinitely many times. This implies that (S_n) will oscillate wildly, with ever larger oscillations, as $n \to \infty$.

Let $S_n = \sum_{i=1}^n X_i$ where the (X_i) are independent and have identical distribution $\mathbb{P}[X_i = 1] = \mathbb{P}[X_1 = -1] = \frac{1}{2}$. Note that $S_0 = 0$. For $k \in \mathbb{Z}$ let

$$T_k = \inf\{n \geq 0\,;\, S_n = k\}$$

be the hitting time of $k \in \mathbb{N}$ and let

$$R = \inf\{n \geq 1\,;\, S_n = 0\}$$

be the first return time to zero. We set $R = \infty$ if the walk does not return to zero, in keeping with the convention that $\inf \emptyset = \infty$. In Example 8.2.2 we showed that T_k is a stopping time, and in Exercise **8.5** we showed that R is a stopping time.

Lemma 9.3.1 *It holds that $\mathbb{P}[R < \infty] = 1$.*

PROOF: Note that $\mathbb{P}[R < \infty] \in [0, 1]$. We will argue by contradiction. Suppose that $\mathbb{P}[R < \infty] = p < 1$. By the strong Markov property, the process $(S_{R+n})_{n\geq 0}$ has the same distribution as (S_n). In words, each time the random walk returns to zero the process restarts, and from then on acts like a simple symmetric random walk, independent of the past. Let $G = \sum_{n=1}^\infty \mathbb{1}\{S_n = 0\}$, which in words counts the number of times (S_n) returns to zero. By the strong Markov property, applied repeatedly at each return, G has a geometric distribution distribution $\mathbb{P}[G = j] = p^j(1-p)$ for all $j \in \{0, 1, \ldots, \}$. In particular, $\mathbb{E}[G] < \infty$.

We will now calculate $\mathbb{E}[G]$ using a different method. Note that $G = \sum_{n=1}^\infty \mathbb{1}\{S_n = 0\}$, hence

$$\mathbb{E}[G] = \sum_{n=1}^\infty \mathbb{E}[\mathbb{1}\{S_n = 0\}] = \sum_{n=1}^\infty \mathbb{P}[S_n = 0]. \tag{9.8}$$

In the first equality above we have used Exercise **6.8** to swap the infinite sum and expectation. Calculating $\mathbb{P}[S_n = 0]$ is Exercise **7.11**; it is zero when n is odd, which leaves us only to consider

$$\mathbb{P}[S_{2n} = 0] \;=\; \binom{2n}{n}2^{-2n} \;=\; \frac{(2n)!}{(n!)^2}2^{-2n} \;\sim\; \frac{\sqrt{4\pi n}(2n/e)^{2n}}{2\pi n(n/e)^{2n}}2^{-2n} \;=\; \frac{1}{\sqrt{\pi n}}. \tag{9.9}$$

In the above we have used Stirling's approximation, in particular equation (9.7), to simplify the binomial coefficients as $n \to \infty$. From (9.9) we obtain that $\sqrt{\pi n}\mathbb{P}[S_{2n} = 0] \to 1$ as $n \to \infty$. Hence there exists some $N \in \mathbb{N}$ such that for all $n \geq N$ we have $\mathbb{P}[S_{2n} = 0] \geq \frac{1}{2\sqrt{\pi n}}$. Putting this back into (9.8) we have

$$\mathbb{E}[G] \geq \sum_{n=N}^\infty \mathbb{P}[S_{2n} = 0] \geq \sum_{n=N}^\infty \frac{1}{2\sqrt{\pi n}} = \infty.$$

We had deduced above that $\mathbb{E}[G]$ was finite, so we have reached a contradiction. Hence in fact $\mathbb{P}[R < \infty] = 1$. ∎

Lemma 9.3.2 *For all $k \in \mathbb{N}$ we have that $\mathbb{P}[T_k < \infty] = 1$.*

PROOF: Note that $T_0 = 0$, because $S_0 = 0$. We next consider T_1. From Lemma 9.3.1 we have $\mathbb{P}[R < \infty] = 1$, hence by the strong Markov property (applied at successive returns to zero) we have that $(S_n)_{n=0}^{\infty}$ visits zero infinitely many times.

At after each visit to zero, by the strong Markov property, the next move of the walk will be to 1 with probability $\frac{1}{2}$, and to -1 with probability $\frac{1}{2}$. Each time, the move will be independent of the past. Hence, we need to wait a Geometric($\frac{1}{2}$) number of times to see a movement to 1. Since this only requires finitely many attempts, we have $\mathbb{P}[T_1 < \infty] = 1$.

Let us now consider T_k for $k \in \mathbb{N}$. To reach k we need to move from 0 to 1, then from 1 to 2, and so on until k. By the strong Markov property applied successively at T_1, then T_2, then T_3 and so on, each such move takes the same number of steps (in distribution) as an independent copy of T_1. Hence the distribution of T_k is that of k i.i.d. copies of T_1. Since $\mathbb{P}[T_1 < \infty] = 1$ it follows immediately that $\mathbb{P}[T_k < \infty] = 1$.

Lastly, we consider $k < 0$. By symmetry of the walk about zero, T_k and T_{-k} have the same distribution, so $\mathbb{P}[T_{-k} < \infty] = \mathbb{P}[T_k < \infty] = 1$. ■

Theorem 9.3.3 *Almost surely, for all $k \in \mathbb{Z}$ there are infinitely many $n \in \mathbb{N}$ such that $S_n = k$.*

PROOF: Since there are only countably many $k \in \mathbb{Z}$, it suffices to prove the result for some fixed (but arbitrary) $k \in \mathbb{Z}$. From Lemma 9.3.2 we have $\mathbb{P}[T_k < \infty] = 1$, so almost surely the walk reaches k, in fact $S_{T_k} = k$. By the strong Markov property, after time T_k the walk continues to behave like a simple symmetric random walk, independent of its past. Lemma 9.3.1 gives that $\mathbb{P}[R < \infty] = 1$, meaning that the walk will almost surely return (again) to k. Repeating this argument, we obtain that almost surely the walk visits k infinitely many times. ■

How long does the simple symmetric random walk take to reach 1?

Let us focus for a moment on the time T_1. Even though we now know that $\mathbb{P}[T_1 < \infty] = 1$, we are not able to apply the optional stopping theorem to (S_n) and T_1, because (S_n) is unbounded and we do not know if $\mathbb{E}[T_1] < \infty$. In fact, in this case $\mathbb{E}[T_1] = \infty$ and the optional stopping theorem does not apply. Instead, we can argue by contradiction and use the optional stopping theorem to *deduce* that $\mathbb{E}[T_1] = \infty$, as follows.

Lemma 9.3.4 *It holds that* $\mathbb{E}[T_1] = \infty$.

PROOF: Suppose that $\mathbb{E}[T_1] < \infty$. Then we could apply the optional stopping theorem to (S_n) and T using the condition (c). This would give $\mathbb{E}[S_0] = 0 = \mathbb{E}[S_{T_1}]$. But, from Lemma 9.3.2 we have $\mathbb{P}[T_1 < \infty] = 1$ which means, by definition of T_1, that $S_{T_1} = 1$ almost surely. This is a contradiction. ∎

Remark 9.3.5 More generally, we have to be very careful about using the optional stopping theorem. The conditions fail in many situations, and there are many examples of stopping times T and martingales (M_n) for which $\mathbb{E}[M_T] \neq \mathbb{E}[M_0]$.

From Lemmas 9.3.2 and 9.3.4 we have that $\mathbb{P}[T_1 < \infty = 1]$ but $\mathbb{E}[T_1] = \infty$. This tells us that although the time T_1 is always finite, it can be very large. You might find this counter-intuitive at first glance since it is clear that $\mathbb{P}[T_1 = 1] = \frac{1}{2}$ by just considering the first step, but the point is that once the walk has (by chance) made a few steps downwards, it starts to take a very long time to find its way back up again. In the next lemma, with a careful calculation we will find out the exact distribution of T_1. Note that T_1 is odd, as the value of (S_n) is odd/even depending on whether n is odd/even.

Lemma 9.3.6 *It holds that* $\mathbb{P}[T_1 = 2n-1] = \binom{2n}{n}2^{-2n}\frac{1}{2n-1} \sim \frac{1}{2n\sqrt{\pi n}}$.

PROOF: For $n = 1$ we have $\mathbb{P}[T_1 = 1] = \mathbb{P}[S_1 = 1] = \frac{1}{2}$, as required. It remains to consider $n \geq 2$. Note that

$$
\begin{aligned}
\mathbb{P}[S_{2n} > 0] &= \mathbb{P}[T_1 \leq 2n-1, S_{2n} > 0] \\
&= \mathbb{E}\left[\mathbb{1}_{\{T_1 \leq 2n-1\}}\mathbb{1}_{\{S_{2n}>0\}}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{\{T_1 \leq 2n-1\}}\mathbb{1}_{\{S_{2n}>0\}} \,|\, \mathcal{F}_{T_1}\right]\right] \\
&= \mathbb{E}\left[\mathbb{1}_{\{T_1 \leq 2n-1\}}\mathbb{E}\left[\mathbb{1}_{\{S_{2n}-S_{T_1}>0\}} \,|\, \mathcal{F}_{T_1}\right]\right] \\
&= \mathbb{E}\left[\mathbb{1}_{\{T_1 \leq 2n-1\}} \times \frac{1}{2}\right] \\
&= \frac{1}{2}\mathbb{P}[T_1 \leq 2n-1].
\end{aligned} \tag{9.10}
$$

In the above, the first line follows because if $S_{2n} > 0$ then $T_1 \leq 2n-1$. The second and third lines use the relationship between expectation and conditional expectation, and the fourth line uses that T_1 is a stopping time, hence $\{T_1 \leq 2n-1\} \in \mathcal{F}_{T_1}$. The fifth line uses the strong Markov property, which gives that $S_{2n} - S_{T_1}$ is independent of $\mathcal{F}_{T_1}$ and behaves like a random walk with $2n - T_1$ steps. Since T_1 is odd, this is an odd number of steps, which means that $S_{2n} - S_{T_1}$ takes odd values and has a distribution that is symmetric about 0, hence $\mathbb{E}[\mathbb{1}_{\{S_{2n}-S_{T_1}>0\}} \,|\, \mathcal{F}_{T_1}] = \frac{1}{2}$.

From the above calculation, and using the symmetry of the random walk about zero, we have that

$$\mathbb{P}[T \le 2n-1] = 2\mathbb{P}[S_{2n} > 0] = \mathbb{P}[S_{2n} > 0] + \mathbb{P}[S_{2n} < 0] = 1 - \mathbb{P}[S_{2n} = 0]. \tag{9.11}$$

To finish, we note that for $n \ge 2$,

$$\begin{aligned}
\mathbb{P}[T = 2n-1] &= \mathbb{P}[T \le 2n-1] - \mathbb{P}[T \le 2n-3] \\
&= (1 - \mathbb{P}[S_{2n} = 0]) - (1 - \mathbb{P}[S_{2n-2} = 0]) \\
&= \binom{2n-2}{n-1} 2^{-2n+2} - \binom{2n}{n} 2^{-2n} \\
&= \binom{2n}{n} 2^{-2n} \frac{1}{2n-1}.
\end{aligned}$$

The second line of the above follows from (9.11), and the last line follows from a short calculation that is left for you. The last claim of the lemma follows from the same calculation as in (9.9). ■

The calculation that led to (9.10) is known as a 'reflection principle'. It applies the strong Markov property at T_k (with $k = 1$, in this case), and then divides the future into two cases based on whether the first move is upwards of downwards. The distribution of the walk in these two cases is a mirror of one another i.e. a reflection about height k.

9.4 Long term behaviour: asymmetric case (Δ)

We will now move on to study the simple asymmetric random walk. We'll recycle some of our previous notation. Let $S_n = \sum_{i=1}^n$ where the (X_i) are independent and have identical distribution $\mathbb{P}[X_i = 1] = p > \frac{1}{2}$ and $\mathbb{P}[X_1 = -1] = 1 - p < \frac{1}{2}$. The case $p < \frac{1}{2}$ can be handled by considering $(-S_n)$ in place of (S_n). Let $R = \inf\{n \geq 1\,;\, S_n = 0\}$ be the first return time of (S_n) to zero. As before, we set $R = \infty$ if the walk does not return to zero.

Like the symmetric random walk, the asymmetric case sees oscillations, but they are dominated by the drift upwards, so much so that (S_n) might not return to origin even once.

Lemma 9.4.1 *It holds that $\mathbb{P}[R < \infty] < 1$ and $S_n \stackrel{a.s.}{\to} \infty$.*

PROOF: We will first prove that $S_n \stackrel{a.s.}{\to} \infty$. In fact, we have already proved this using a martingale convergence argument in Exercise 7.10. We'll give a different proof here, using the strong law of large numbers. We have $S_n = \sum_{i=1}^n X_i$, where the X_i are i.i.d. random variables, so the strong law of large numbers implies that $\frac{S_n}{n} \stackrel{a.s.}{\to} \mathbb{E}[X_1]$ as $n \to \infty$. Note that $\mathbb{E}[X_1] = p(1) + (1-p)(-1) = 2p - 1$. It follows that there exists (a random variable) $N \in \mathbb{N}$ such that for all $n \geq N$ we have $\frac{S_n}{n} \geq \frac{2p-1}{2}$, which implies $S_n \geq n\frac{2p-1}{2}$. Since $2p - 1 > 0$ we therefore have $S_n \stackrel{a.s.}{\to} \infty$. In particular this means that $\mathbb{P}[(S_n)$ visits zero infinitely many times$] = 0$.

If $\mathbb{P}[R < \infty]$ was equal to 1 then, by applying the strong Markov property at each successive return time to zero, we would have that almost surely (S_n) visited the origin infinitely many times. We have shown above that this is not the case. Hence $\mathbb{P}[R < \infty] < 1$. ∎

Remark 9.4.2 By the strong Markov property and the same argument as in the proof of Lemma 9.3.1 (except that there it was part of a proof by contradiction) the number of times that (S_n) returns to the origin is a Geometric($\hat{p}$) random variable, where $\hat{p} = \mathbb{P}[R < \infty]$.

If we take T_k to be the hitting time of $k \in \mathbb{Z}$, then it follows immediately from Lemma 9.4.1 that $\mathbb{P}[T_k < \infty] = 1$ for $k \geq 0$ and $\mathbb{P}[T_k < \infty] < 1$ for $k < 0$. In the symmetric case Lemma 9.3.4 showed that $\mathbb{E}[T_1] = \infty$, but here in the asymmetric case the opposite is true.

Lemma 9.4.3 *It holds that $\mathbb{E}[T_1] = \frac{p}{2p-1}$.*

PROOF: We calculate

$$\begin{aligned}
\mathbb{E}[T_1] &= \mathbb{E}[T_1 \mathbb{1}_{\{S_1=1\}}] + \mathbb{E}[T_1 \mathbb{1}_{\{S_1=-1\}}] \\
&= \mathbb{E}[1 \mathbb{1}_{\{S_1=1\}}] + \mathbb{E}[\mathbb{E}[T_1 \mathbb{1}_{\{S_1=-1\}} \,|\, \mathcal{F}_1]] \\
&= p + \mathbb{E}[\mathbb{1}_{\{S_1=-1\}} \mathbb{E}[T_1 \,|\, \mathcal{F}_1]] \\
&= p + \mathbb{E}[\mathbb{1}_{\{S_1=-1\}} \mathbb{E}[1 + T_1' + T_1'' \,|\, \mathcal{F}_1]] \\
&= p + \mathbb{E}[\mathbb{1}_{\{S_1=-1\}} \mathbb{E}[1 + T_1' + T_1'']] \\
&= p + \mathbb{E}[\mathbb{1}_{\{S_1=-1\}} (1 + 2\mathbb{E}[T_1])] \\
&= p + (1-p)(1 + 2\mathbb{E}[T_1]).
\end{aligned}$$

In the first line of the above we partition on the value of S_1. The first term of the second line follows because if $S_1 = 1$ then $T_1 = 1$, and the second term uses the relationship between expectation and conditional expectation. The third line follows because $S_1 \in \mathcal{F}_1$. The fourth line uses that, on the event $S_1 = -1$, T_1 is equal to 1 (accounting for the first move $0 \mapsto -1$) plus two independent copies (T_1' and T_1'') of T_1 (accounting for the time to move from $-1 \mapsto 0$ and then $0 \mapsto 1$). The strong Markov property gives that T_1' and T_1'' are independent of $\mathcal{F}_1$, leading to the fifth line. The remainder of the calculation is straightforward. Rearranging to isolate $\mathbb{E}[T_1]$, we obtain the required result. ■

Remark 9.4.4 Using a similar argument to that used within the proof of Lemma 9.3.2, for $k \in \mathbb{N}$ we have that T_k is the sum of k i.i.d. copies of T_1, hence $\mathbb{E}[T_k] = \frac{kp}{2p-1}$.

Note that when $p \approx 1$ we have $\mathbb{E}[T_1] \approx 1$, which makes sense since when p is close to 1 it is very likely that the first step of the walk will be upwards. As $p \searrow \frac{1}{2}$ we have $\mathbb{E}[T_1] \to \infty$, which also makes sense because as p gets close to $\frac{1}{2}$ we would expect the asymmetric walk to look more and more like the symmetric case, which has $\mathbb{E}[T_1] = \infty$.

9.5 In higher dimensions (⊘)

We will look briefly at the behaviour of simple symmetric random walks in higher dimensions. This section is off-syllabus and is included for interest. In particular, we will consider versions of Lemma 9.3.1 in dimension $d = 2$ and $d \geq 3$.

We must first explain what the simple symmetric random walk is, in higher dimensions. It will suffice to describe $d = 2$ to make the principle clear. Then, for each $n \in \mathbb{N}$, S_n is a random element of $\mathbb{Z}^2$, which we write as $S_n = (S_n^{(1)}, S_n^{(2)})$. The movements are given by $S_n = \sum_{i=1}^n X_i$ where $\mathbb{P}[X_i = (0,1)] = \mathbb{P}[X_i = (0,-1)] = \mathbb{P}[X_i = (1,0)] = \mathbb{P}[X_i = (0,-1)] = \frac{1}{4}$. These four possibilities correspond to movements by precisely one unit of space in the four possible directions: north, south, east and west.

Let $R = \inf\{n \geq 1\,;\, S_n = \mathbf{0}\}$ denote the time taken to return to the origin $\mathbf{0} = (0,0)$. In two dimensions we have:

Lemma 9.5.1 *If $d = 2$ then $\mathbb{P}[R < \infty] = 1$.*

PROOF: We may use much the same argument as in Lemma 9.3.1. The only difference is that we must replace the estimate 9.9 with a two dimensional version, as follows. We have

$$\begin{aligned}\mathbb{P}[S_{2n} = (0,0)] &= \sum_{i=0}^{n} \binom{2n}{2i}\binom{2i}{i}\binom{2n-2i}{n-i} 4^{-2n} \\ &= \binom{2n}{n} \sum_{k=0}^{n} \binom{n}{k}^2 4^{-2n} \\ &= \left[\binom{2n}{n} 2^{-2n}\right]^2 \\ &\sim \frac{1}{\pi n}. \end{aligned} \tag{9.12}$$

In the above, the first line comes from counting the number of possible ways in which the two dimensional walk can start at 0 and returns to 0 after exactly $2n$ steps (we did the one dimensional version in Exercise **7.11**). To do so the walk must have made the same number of steps north as it did south, and the same number of steps east as it did west. We write $2i$ for the number of east-west steps and $2j = 2n - 2i$ for the number of north-south steps. The first binomial factor counts which steps are east-west and which are north-south. The second and third, respectively, count which of the east-west steps are to the east, and which of the north-south steps are to the north – and that fixes everything else about the path.

The second line of the above calculation follows easily from the first, and you should write the factorials out in full to check this for yourself. The third line follows from the second by the combinatorical identity $\binom{2n}{n} = \sum_{k=0}^{n} \binom{n}{k}$. The final line follows from the same calculation as in the right hand side of (9.9).

With (9.12) in place of (9.9), the rest of the argument for this lemma follows in the same way as that of Lemma 9.3.1. We'll omit writing out the details again. ∎

We won't prove it here, but the simple symmetric random walk will take *much* longer to return to the origin in $d = 2$ than $d = 1$. This is because in order to reach $(0,0)$ both the east-west component $S_n^{(1)}$ and the north-south component $S_n^{(2)}$ must return simultaneously to zero, and they unlikely to do this at the same time. Perhaps surprisingly, in two dimensions the random walk still manages to return. The picture changes if we move to three dimensions, with the obvious notation: now $S_n \in \mathbb{Z}^3$ and there are 6 different directions to move in, each with probability $\frac{1}{6}$. In three dimensions there is now so much space to get lost in that the random walk might not return to the origin.

Lemma 9.5.2 *If $d = 3$ then $\mathbb{P}[R < \infty] < 1$.*

Sketch of Proof: We will only give a sketch of the argument. We need to make some substantial modifications to the argument used in Lemmas 9.3.1 and 9.5.1, because we are now trying to prove the opposite result, but there are still parts that overlap. In similar style to (9.12) we have

$$\begin{aligned}
\mathbb{P}[S_{2n} = (0,0,0)] &= \sum_{i+j+k=n} \frac{(2n)!}{(i!)^2(j!)^2(k!)^2} 6^{-2n} \\
&= \binom{2n}{n} 2^{-2n} \sum_{i+j+k=n} \left(\frac{n!}{i!j!k!} 3^{-n}\right)^2 \\
&\leq \binom{2n}{n} 2^{-2n} \left(\max_{i+j+k=n} \frac{n!}{i!j!k!} 3^{-n}\right) \left(\sum_{i+j+k=n} \frac{n!}{i!j!k!} 3^{-n}\right).
\end{aligned}$$

The first line again concerns counting possible ways to return to the origin in exactly $2n$ steps (we'll omit the details of this) and the next two lines follow by elementary calculations. We have that $\sum_{i+j+k=n} \frac{n!}{i!j!k!} 3^{-n} = 1$, because this sums up all the probabilities of outcomes (i,j,k) that sum to n according to the trinomial distribution (which you might have to look up, if you haven't seen it before). A bit of calculus with Lagrange multipliers shows that the term $\frac{n!}{i!j!k!} 3^{-n}$, subject to the condition $i + j + k = n$, is maximised when i, j, k are all approximately equal to $\frac{n}{3}$. Approximating $\frac{n}{3}$ by the nearest integer, we can find that $\max_{i+j+k=n} \frac{n!}{i!j!k!} 3^{-n} \leq \frac{C}{n}$, where $C \in (0,\infty)$ is a constant. Combining this with the calculation from (9.9) we obtain that $\mathbb{P}[S_{2n} = \mathbf{0}] \leq \frac{C'}{n^{3/2}}$, where $C' \in (0,\infty)$ is also constant.

Setting $G = \sum_{n=1}^{\infty} \mathbb{1}\{S_n = \mathbf{0}\}$ we now have

$$\mathbb{E}[G] = \sum_{n=1}^{\infty} \mathbb{P}[S_{2n} = \mathbf{0}] \leq \sum_{n=1}^{\infty} \frac{C'}{n^{3/2}} < \infty.$$

Hence the expected number of visits to the origin is finite. By the strong Markov property, if $\mathbb{P}[R < \infty] = 1$ then we would almost surely have infinitely many returns to the origin. Hence $\mathbb{P}[R < \infty] < 1$. ■

Lemma 9.5.2 also holds in all dimensions $d \geq 3$, To see this, note that the first three coordinates of a d-dimensional random walk give us a three dimensional random walk. If the d-dimensional walk visits the origin, so do its first three coordinates, but Lemma 9.5.2 gives that this has probability less than one.

It is natural to ask what the random walk does do, in three dimensions and higher, as $n \to \infty$. The answer is easy to guess: it disappears off to infinity and does not come back i.e. $|S_n| \stackrel{a.s.}{\to} \infty$. This is left for you to prove, as Exercise **9.9**.

9.6 Exercises on Chapter 9 (Δ)

In all questions below (S_n) denotes a random walk started at the origin, but which random walk varies according to the question.

On one dimensional random walks

9.1 Let (S_n) be the simple asymmetric random walk, as in Section 9.1. Let $a < 0 < b$ be integers and define the hitting times $T_a = \inf\{n \in \mathbb{N}\,;\, S_n = a\}$, $T_b = \inf\{n \in \mathbb{N}\,;\, S_n = b\}$ and $T = T_a \wedge T_b$.

(a) Show that $\mathbb{E}[S_T] = (p - q)\mathbb{E}[T]$.

(b) Calculate $\mathbb{E}[S_T]$ directly using (9.5) and (9.6) and hence calculate $\mathbb{E}[T]$.

9.2 *This question applies some of the techniques from Section 9.1 to the symmetric case.*

Let (S_n) denote the simple symmetric random walk and let T_k be the hitting time of $k \in \mathbb{Z}$. Let $a < 0 < b$ be integers and let $T = T_a \wedge T_b$.

(a) Explain carefully why both

$$\begin{aligned} 1 &= \mathbb{P}[T = T_a] + \mathbb{P}[T = T_b] \\ 0 &= a\mathbb{P}[T = T_a] + b\mathbb{P}[T = T_b]. \end{aligned}$$

Hint: Recall that (S_n) is a martingale.

(b) Solve these equations to find explicit formulae for $\mathbb{P}[T = T_a]$ and $\mathbb{P}[T = T_b]$ in terms of a and b.

(c) Show that $\mathbb{E}[T] = -ab$.

Hint: Can you think of a useful martingale?

9.3 Let (X_i) be a sequence of independent, identically distributed random variables with $\mathbb{P}[X_i = 2] = \frac{1}{3}$ and $\mathbb{P}[X_i = -1] = \frac{2}{3}$. Set

$$S_n = \sum_{i=1}^{n} X_i$$

and define the stopping time $R = \inf\{n \geq 1\,;\, S_n = 0\}$.

(a) Modify the argument in Exercise **7.11** to calculate $\mathbb{P}[S_n = 0]$ explicitly. Hence show that $\mathbb{P}[S_{3n} = 0] \sim \frac{\sqrt{3}}{2\sqrt{\pi n}}$ as $n \to \infty$.

(b) Explain how to modify the proof of Lemma 9.3.1 to deduce that $\mathbb{P}[R < \infty] = 1$.

9.4 Let $p \in (\frac{3}{5}, 1]$. Let (X_i) be a sequence of independent, identically distributed random variables with $\mathbb{P}[X_i = 1] = p$ and $\mathbb{P}[X_i = -1] = \mathbb{P}[X_i = -2] = \frac{1-p}{2}$. Set

$$S_n = \sum_{i=1}^{n} X_i$$

and let $T_1 = \inf\{n \geq 1\,;\, S_n = 1\}$.

(a) Explain how to modify the argument in Lemma 9.4.1 to show that $S_n \stackrel{a.s.}{\to} \infty$ as $n \to \infty$. Hence show that $\mathbb{P}[T_1 < \infty] = 1$.

(b) Modify the argument in Lemma 9.4.3 to calculate $\mathbb{E}[T_1]$.

9.5 For the simple symmetric random walk, in Lemma 9.3.6 we showed that $\mathbb{P}[T_1 = 2n-1] \sim \frac{1}{2n\sqrt{\pi n}}$. Use this fact to give a second proof (alongside that of Lemma 9.3.4) that $\mathbb{E}[T_1] = \infty$.

9.6 Let (S_n) denote the simple symmetric random walk and let $T_m = \inf\{n \geq 0\,;\, S_n = m\}$ be the first hitting time of $m \in \mathbb{Z}$. Let

$$M_n^{(\theta)} = \frac{e^{\theta S_n}}{(\cosh\theta)^n}$$

where $\theta \in \mathbb{R}$.

(a) Show that $M_n^{(\theta)}$ is a martingale.

(b) Check that none of the conditions (a)-(c) of the optional stopping theorem apply to the martingale $(M_n^{(\theta)})$ at the stopping time T_m.

(c) **[Challenge question]** Show that

$$\mathbb{E}\left[\frac{1}{(\cosh\theta)^T}\right] = \frac{1}{\cosh(m\theta)}$$

where $T = T_m \wedge T_{-m}$. You should start by applying the optional stopping theorem to a suitable martingale.

On random walks in two and three dimensions $(\oslash)$

9.7 Let (S_n) denote the two dimensional simple symmetric random walk, as defined in Section 9.5. Prove that, almost surely, for each $z \in \mathbb{Z}^2$ there are infinitely many $n \in \mathbb{N}$ such that $S_n = z$.

Hint: You can re-use some of the ideas from proof of Theorem 9.3.3.

9.8 Let (S_n) denote the three dimensional simple symmetric random walk, as defined in Section 9.5. Let $G = \sum_{n=0}^{\infty} \mathbb{1}_{\{S_{2n}=0\}}$ denote the total number of visits to the origin. Let $R = \min\{n = 1, 2, \ldots\,;\, S_n = 0\}$ and $L = \max\{n = 0, 1, 2, \ldots\,;\, S_n = 0\}$ denote, respectively, the first return time and the last visiting time of (S_n) to the origin.

(a) Explain why $\mathbb{P}[L < \infty] = 1$, as a consequence of Lemma 9.5.2.

(b) Is L is a stopping time? Give a brief reason for your answer.

(c) Show that $\mathbb{P}[L = 2n] = \mathbb{P}[S_{2n} = 0]\mathbb{P}[R = \infty]$ and hence prove that $\mathbb{E}[G] = \frac{1}{1-\mathbb{P}[R<\infty]}$.

(d) We already came close to deducing this exact formula for $\mathbb{E}[G]$, more than once, within the current chapter. Can you see where?

9.9 Let (S_n) denote the three dimensional simple symmetric random walk, as defined in Section 9.5. Prove that $|S_n| \stackrel{a.s.}{\to} \infty$ as $n \to \infty$.

Appendix A

Solutions to exercises (part one)

Solutions are omitted from the printed lecture notes. You can find them inside the online version of these notes.

Appendix B

Formula Sheet (part one)

The formula sheet displayed on the following page will be provided in the exam.

MAS352/61023 – Formula Sheet – Part One

Where not explicitly specified, the notation used matches that within the typed lecture notes.

Modes of convergence

- $X_n \stackrel{d}{\to} X \;\Leftrightarrow\; \lim_{n\to\infty} \mathbb{P}[X_n \le x] = \mathbb{P}[X \le x]$ whenever $\mathbb{P}[X \le x]$ is continuous at $x \in \mathbb{R}$.
- $X_n \stackrel{\mathbb{P}}{\to} X \;\Leftrightarrow\; \lim_{n\to\infty} \mathbb{P}[|X_n - X| > a] = 0$ for every $a > 0$.
- $X_n \stackrel{a.s.}{\to} X \;\Leftrightarrow\; \mathbb{P}[X_n \to X \text{ as } n \to \infty] = 1$.
- $X_n \stackrel{L^p}{\to} X \;\Leftrightarrow\; \mathbb{E}[|X_n - X|^p] \to 0$ as $n \to \infty$.

The binomial model and the one-period model

The binomial model is parametrized by the deterministic constants r (discrete interest rate), p_u and p_d (probabilities of stock price increase/decrease), u and d (factors of stock price increase/decrease), and s (initial stock price).

The value of x in cash, held at time t, will become $x(1+r)$ at time $t+1$.

The value of a unit of stock S_t, at time t, satisfies $S_{t+1} = Z_t S_t$, where $\mathbb{P}[Z_t = u] = p_u$ and $\mathbb{P}[Z_t = d] = p_d$, with initial value $S_0 = s$.

When $d < 1 + r < u$, the risk-neutral probabilities are given by

$$q_u = \frac{(1+r) - d}{u - d}, \qquad q_d = \frac{u - (1+r)}{u - d}.$$

The binomial model has discrete time $t = 0, 1, 2, \ldots, T$. The case $T = 1$ is known as the one-period model.

Stirling's Approximation

It holds that $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$.

Conditions for the optional stopping theorem (MAS61023 only)

The optional stopping theorem, for a martingale M_n and a stopping time T, holds if any one of the following conditions is fulfilled:

(a) T is bounded.

(b) $\mathbb{P}[T < \infty] = 1$ and there exists $c \in \mathbb{R}$ such that $|M_n| \le c$ for all n.

(c) $\mathbb{E}[T] < \infty$ and there exists $c \in \mathbb{R}$ such that $|M_n - M_{n-1}| \le c$ for all n.